Master's thesis
Degree Programme in Particle physics and Astrophysics
Study track Astrophysics

SPIDERS - The making of an X-ray detected spectroscopic galaxy cluster survey and catalogue

Sanna Damsted

19th of May 2020

Supervisor(s): Alexis Finoguenov

Examiner(s): Charles C. Kirpatrick IV

UNIVERSITY OF HELSINKI
FACULTY OF SCIENCE

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Laitos — Institution — Department
Faculty of Science	Department of physics

Tekijä — Författare — Author
Sanna Damsted

Työn nimi — Arbetets titel — Title

SPIDERS - The making of an X-ray detected spectroscopic galaxy cluster survey and catalogue

Oppiaine — Läroämne — Subject
Astrophysics

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's thesis	May 19, 2020	65 pages + 0 appendix pages

Tiivistelmä — Referat — Abstract

Galaxy clusters are the largest tightly gravitationally bound structures in the Universe. They are abundant and uniformly distributed in the sky, which makes them excellent targets for studying the history and properties of our Universe. They can be used to study some of the burning questions in astrophysics and cosmology. For example, how have the large scale structures evolved from billions of years ago to how we see them today, and what are the best values for parameters like matter densities in our cosmological theories.

Photometric galaxy cluster surveys have been conducted for decades, but now is the dawn of large scale spectroscopic surveys. Spectroscopy gives more precise redshift values for galaxies and galaxy clusters than photometric observations, which in turn makes other astrophysical parameters like mass estimation of galaxy clusters much more reliable.

The SPIDERS galaxy cluster survey is the largest X-ray detected, spectroscopic, visually validated survey conducted to date. It improves the precision of galaxy cluster redshifts by a factor of 10. The precision of redshift has a direct improvement on other distance related parameters calculated by using galaxy clusters. The SPIDERS value added catalogue, which came out of the survey, is a tremendous achievement and will benefit astrophysicists and cosmologist around the world. The catalogue is the result of the work of the SPIDERS team of experts, and it is freely available online.

This thesis explains how the SPIDERS survey was conducted; it's phases, algorithms and the science behind it. I give many examples of the data processing and visual validation of targets, and explain the results and the significance of having such a large and precise data set.

Avainsanat — Nyckelord — Keywords
Galaxy clusters, Galaxy cluster survey, X-ray astronomy, Optical astronomy, Spectroscopic observations

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — övriga uppgifter — Additional information

Contents

1 Motivation for the thesis

Galaxy clusters are laboratories for studying processes operating during galaxy formation and their effects on the surrounding intergalactic medium. They are abundant and uniformly distributed across the sky so they can be used for studying cosmological properties and the history of the Universe. ΛCDM theory predicts that the abundance of clusters is determined by the spectrum of primordial perturbations and cosmological parameters λ and Ω. To study these parameters with precision, large surveys with well defined mass functions are needed. Galaxy clusters can be used to study the overall mass density of the Universe and the amplitude of the initial spectrum of density perturbations in the early Universe. [1][2]

There is a tension between the reported values of amplitude of spectral fluctuations in the present day (σ_8) between different types of observations. Cosmic microwave background surveys (Planck and WMAP) infer a higher value for σ_8 than rich galaxy cluster counts (SZ) and weak lensing surveys. [3] Studies of mass estimations of galaxy clusters using different methods have found tension between estimates using ΛCDM prior obtained from the cosmic microwave background (CMB) to other methods. In a study of dynamical mass estimates from Sunyaev-Zel'dovich (SZ) selected massive galaxy clusters, that compared to a strong cosmological prior, mass estimates were 81% of the compared values. These mass estimates were consistent with other studies done using velocity dispersions, weak lensing and Y_X masses derived from measured hot gas distribution in X-ray. [4]

Spectroscopic surveys are the best options for studying dynamical properties of galaxy clusters in the optical spectrum. They are also essential for estimating many other properties like X-ray masses or cosmological parameters, which need precise distance measurements. Photometric surveys are good, but tend to suffer from projection effects, because they cannot accurately measure the velocity space of the galaxies, only their colors. [5] Also at higher redshifts, photometric observations have large scatter that will affect precision. We are witnessing a time where large photometric surveys are being done such as SDSS, PS1 and the future LSST survey. Large spectroscopic surveys are slower to conduct, but surveys such as SDSS and future 4MOST and WEAVE will change how astronomy is done. Validation is also a major part of creating reliable samples and should be done for any creditable catalogue. Large surveys are needed for statistically meaningful samples. In the past, spectroscopic cluster samples have had some tens or a couple hundred clusters, but this has now been changed with the SPIDERS (SPectroscopic IDentification of eROSITA Sources) spectroscopic catalogue, which is the main focus of this thesis. The final SPIDERS catalogue consist of over 2700 validated clusters in the redshift range of $0.01 < z < 0.7$. The SPIDERS catalogue can be used to answer some of the burning questions in modern astrophysics and observational cosmology.

This thesis is constructed as such: in section 2 I introduce the background and theoretical aspects needed to understand the thesis, in section 3 I explain the data and methods of handling the data, in section 4 I explain what is done to the data within the SPIDERS program and a short description of another project with the

same data: substructure in the galaxy clusters, in section 5 I present the results, in section 6 I discuss the meaningfulness of the results and the importance of having a large spectroscopic catalogue of galaxy clusters and the conclusions of the thesis.

2 Theory

2.1 Cosmological background

2.1.1 ΛCDM

The ΛCDM paradigm is a widely supported theory of the evolution and structure of our Universe. It answers to many questions which have baffled astrophysicists and cosmologists for decades. There are few basic concepts, which describe our Universe. First, we live in a Universe which is "flat" meaning that the geometry of the Universe has no curvature. Everything started with a Big bang. The Universe consists of baryonic (normal) matter; radiation; collisionless, "cold" dark matter (CDM) and some form of repulsive force (Λ) that is responsible for the accelerating expansion of the Universe. Cold dark matter refers to free streaming length of dark matter particles being smaller than a protogalaxy, which leads to a hierarchical building of large scale structures. [6]

The five cosmological parameters of the ΛCDM model are: the physical baryon density $\Omega_b h^2$, the physical cold dark matter density $\Omega_c h^2$, the dark energy density in units of the critical density Ω_Λ, the amplitude of primordial scalar curvature perturbations Δ_2^R at $k = 0.002 Mpc^{-1}$ and the power-law spectral index of primordial density (scalar) perturbations n_s. [7] In a ΛCDM Universe, baryonic matter or "normal" matter is believed to have a density of 4.6% of critical density and a dark matter density of 23.6% of critical density. The rest is dark energy. Critical density is defined as $\rho_{cr} = 3H_0^2/8\pi G$, where H_0 is the Hubble constant at current time and G is the gravitational constant. [6][5]

2.1.2 Other theories

ΛCDM isn't the only theory that may be able to describe what we see when we are looking at the evolution and observational properties of the Universe. Modified gravity theories have been developed, where gravity behaves differently at different scales. These theories have their supporters in the science community and should not be overlooked, although they are not the favored theories of the majority. This thesis does not take sides on the matter of preferred cosmological theory since it is an observationally motivated study. However, galaxy cluster research is an important tool that may be used to test different cosmological theories.

2.2 Galaxy clusters

Galaxy clusters are the biggest tightly gravitationally bound systems in the Universe. They contain from a few hundred to thousands of galaxies and have masses from 10^{14} to 10^{15} solar masses. Stars and other galactic matter contribute roughly a tenth of the total baryonic mass and very hot $10^7 K$ intracluster gas contributes most of the rest of the baryonic mass in galaxy clusters. The majority of mass in galaxy clusters is dark matter according to the standard cosmological model. The formation of galaxy clusters is thought to happen hierarchically in the current consensus under ΛCDM theory. Clusters merge together from smaller structures to larger ones and are driven together by the underlying gravitational force. The term for this is "growth of structure" and it plays an important role in understanding the history - and future - of our Universe. [1] Observations of galaxy clusters have been made in a whole range of wavelengths using different methods. In this thesis, the focus is on two: optical and X-rays. A short description of these and some other relevant mass estimation methods is given in the following subsections.

Galaxy clusters and groups have a whole range of different galaxies in them, but not evenly distributed in them. Our local Universe consists mainly of irregular dwarf galaxies and spiral galaxies. Bigger clusters also have many dwarf galaxies and spiral galaxies at, but the majority of the stellar mass is observed in elliptical galaxies. The difference between groups and clusters comes from their evolution.

The λCDM models predict that large clusters have larger dark matter subhalos in them than groups. This will result in larger galaxies in them. The difference in galaxy types comes also from the hierarchical model where galaxy groups come together and make larger groups and clusters. In the process, galaxies interact with each other and with the group's potential well. This sets up bursts of star formation due to infalling gas, which cools while it gets denser and that can then turn into proto-stars if they reach a critical mass. In rich clusters however, when merging clumps of galaxies start falling towards the core of the cluster, gas in the galaxies and between the galaxies starts colliding with the vast amount of hot intracluster gas in the cluster's inner regions. This stripping prevents new star formation and only the current stars in the galaxies are left. Therefore stars in the cluster galaxies tend to be older than in groups or at the outskirts of clusters. The age of stars correlates with their colors, so simplified: in the core of clusters, galaxies are mainly red galaxies. [8]

At the core of galaxy clusters usually also lies a large elliptical galaxy. At the bottom of the potential well of the cluster, galaxies are more likely to merge with each other than at the outskirts of clusters. This results in the largest central galaxy to grow faster than other galaxies in the cluster. As mass and brightness of galaxies correlates, these central galaxies are also brighter than the rest of the cluster galaxies and they are hence called brightest cluster galaxies (BCG).

BCGs are generally identified from photometric observations by taking the brightest source matching the color of the cluster. More reliable identification can be done

with spectroscopic observations, since it gives a more precise velocity offset from the mean of the cluster. It's not uncommon to misidentify BCGs in photometric observations. They don't always lie in the dynamical center of the cluster and in some cases there may be more than one galaxy of similar brightness in the same cluster. [9] Also using a cluster finder that's based on colors of the galaxies, like a red sequence finder (see section 3.3.2), the BCG may not be found, if it is a cool-core galaxy and is having major star formation, which makes it bluer than the majority of the cluster galaxies.

To get information about galaxy clusters and large scale structure, observations are needed. For doing any statistical analysis based on observations, conducting surveys is essential. For the purpose of testing cosmological theories with the help of galaxy clusters, large numbers are a requirement, the larger the better. Estimating properties such as X-ray luminosities, or masses for X-ray or Sunyaev-Zel'dovich detected clusters, the distances to the object must be known. The same is true for any cosmological properties regarding dark matter, dark energy or the evolution of large scale structures: we need distances to the objects. Measuring redshifts is the way to do this. Using photometric observations may be sufficient, but to do precise estimations, spectroscopic observations are crucial. Large scale spectroscopic galaxy cluster surveys are the future of galaxy cluster astrophysics and precision cosmology.

2.3 Mass estimation of galaxy clusters

There is no direct way to measure masses of galaxy clusters. Many different means can however be used to estimate masses. Most common mass proxies are briefly described in this section.

2.3.1 X-ray properties

Galaxy clusters have deep gravitational potential wells. The mass in clusters is mainly dark matter, but the majority of the baryonic matter in galaxy clusters is in the form of intracluster medium (ICM). ICM is plasma which consists mainly of ionized Hydrogen and Helium. [8] Plasma in a gravitational potential well produces thermal Bremsstrahlung - breaking radiation. Bremsstrahlung is the deceleration of particles where the excess energy is emitted away as photons. At $10^7 K$ temperatures the emitted photons are X-ray photons according to $k_B T = h\nu$, where k_B is the Boltzmann constant, T is the temperature, h is the Planck constant and ν is the frequency. [10] These hot clumps of gas are observable with modern X-ray telescopes. In fact they are very prominent in X-ray surveys and are a valid method for detecting galaxy clusters and also estimating the underlying mass of the clusters. X-ray observations can be used to measure the ICM mass and temperature and from those the total mass of the clusters can be estimated. This requires however, that distances to the objects are known. Getting enough X-ray photons from thousands of targets to measure their redshifts is not feasible with current equipment available, so optical observations are needed for determining distances.

2.3.2 Dynamical properties

Observations have shown that the velocities of the member galaxies and the temperature of the intracluster gas are correlated. The intracluster gas is not associated directly to the galaxies, but both baryonic components are indicators of the common potential well. [1]

Objects in a common gravitational potential will trace the mass of the system with their radial velocities and positions in the system. Velocity dispersion about the mean of the velocity of the group of objects can be used as the method for mass estimation of the system. This is called the dynamical mass of the system. It follows from the virial theorem, which states that the total kinetic energy of the system can be calculated from the radii and forces of the particles in that system. The base of the theorem is in classical mechanics and has been around for a couple of centuries by now.

Velocity dispersion is an excellent property to use for testing for example cosmological theories, because it is local and does not depend on cosmological parameters used. Even if some parameter changes, velocity dispersion is unaffected.

In the classical scheme the system is assumed spherical and isolated, which is a rather simplified statement for galaxy clusters. Implementing velocity dispersion to galaxy clusters works in principle if the system is "relaxed" ie. the velocity distribution of member galaxies follows a Gaussian distribution. This is mostly not the case. Galaxy clusters are not usually completely relaxed; they may have heavier than normal Gaussian tails; have substructure; may contain galaxies that are still on their first orbit in the system or they can be interacting or merging with other clusters. This all means there is a need for some other type of estimator. [11]

Better estimators have been developed for velocity dispersion for astronomical purposes. For example, the bi-weight estimator and the gapper estimator, which use different center location and scale compared to the mean and dispersion of the classical velocity dispersion. A more detailed description of these methods can be found in section 3.6.

2.3.3 Richness

Richness is an important attribute of galaxy clusters and it is used widely in astrophysics and cosmological research. Richness is essentially the number of galaxies in a galaxy cluster. But we can't detect all the galaxies even in our local Universe, so other definitions must be used. Abell was the first to construct a proper galaxy cluster catalogue with a richness measurement in 1958. His definition was to count all galaxies within 2 magnitudes from the third brightest galaxy inside an area with radius of $1.7'/z$, where z is the redshift. [12] Other methods have been developed over the years. A well defined iterative richness estimation called redMaPPer richness - which is used as the richness estimate in SPIDERS also - is introduced in section 3.3.2. In estimating richness, many things must be taken into consideration

such as accuracy of redshift measurement, finding a correct center, understanding foreground/background galaxy density and projection effects.

The value of estimating richnesses of galaxy clusters is that it's an observational measurement which is related tightly to the mass of the cluster. Knowing the mass-richness relation of galaxy clusters is one of the means to constrain cosmological parameters at different epochs of time.

2.3.4 Caustic mass estimation

Caustics mean the in-fall patterns of galaxies in galaxy clusters. These patterns are present in almost all rich X-ray luminous clusters at low and moderate redshifts. The identifiable trumpet shape is visible in redshift-radius diagrams. Figure 1 has an example of ensemble clusters where multiple clusters have been stacked together. The solid lines show the caustic shapes and the galaxies inside are almost all cluster member galaxies. Galaxies outside the caustics have escape velocities likely too high to stay in the clusters for extended periods. The total bound masses for the clusters are estimated using the escape velocities and assuming spherical symmetries of the clusters. To reliably estimate the caustic masses, some hundreds of spectra are needed. [13]

2.3.5 Weak lensing

In Einstein's theory of general relativity, light and time are affected by gravity. A massive object has a massive gravitational potential which can bend and magnify the light coming from behind it. This effect is called gravitational lensing and it happens on all scales if we can just get precise enough observations. Single massive objects like black holes can create strong gravitational lensing or even rings around the object. In galaxy cluster astrophysics, lenses are more often massive galaxies or clusters instead of single objects, where the effects are much more subtle. This is called weak lensing. What is measured in weak lensing studies is shear. Shear is the difference of compression or expansion in radial versus tangential directions in the image that we see. In other words the surface brightness does not change, but the area that the image of the source takes on the sky is different depending on the shear. In practise, weak lensing surveys study the shapes of galaxies compared to their neighbours and estimate statistically the masses and shapes of the lenses. Mostly weak lensing masses correlate well with X-ray observations, but there is scatter from non-relaxed clusters, where intracluster gas is heated up by the colliding gas particles, and also from projection effects, where there may be more than one galaxy cluster acting as a lens. [8]

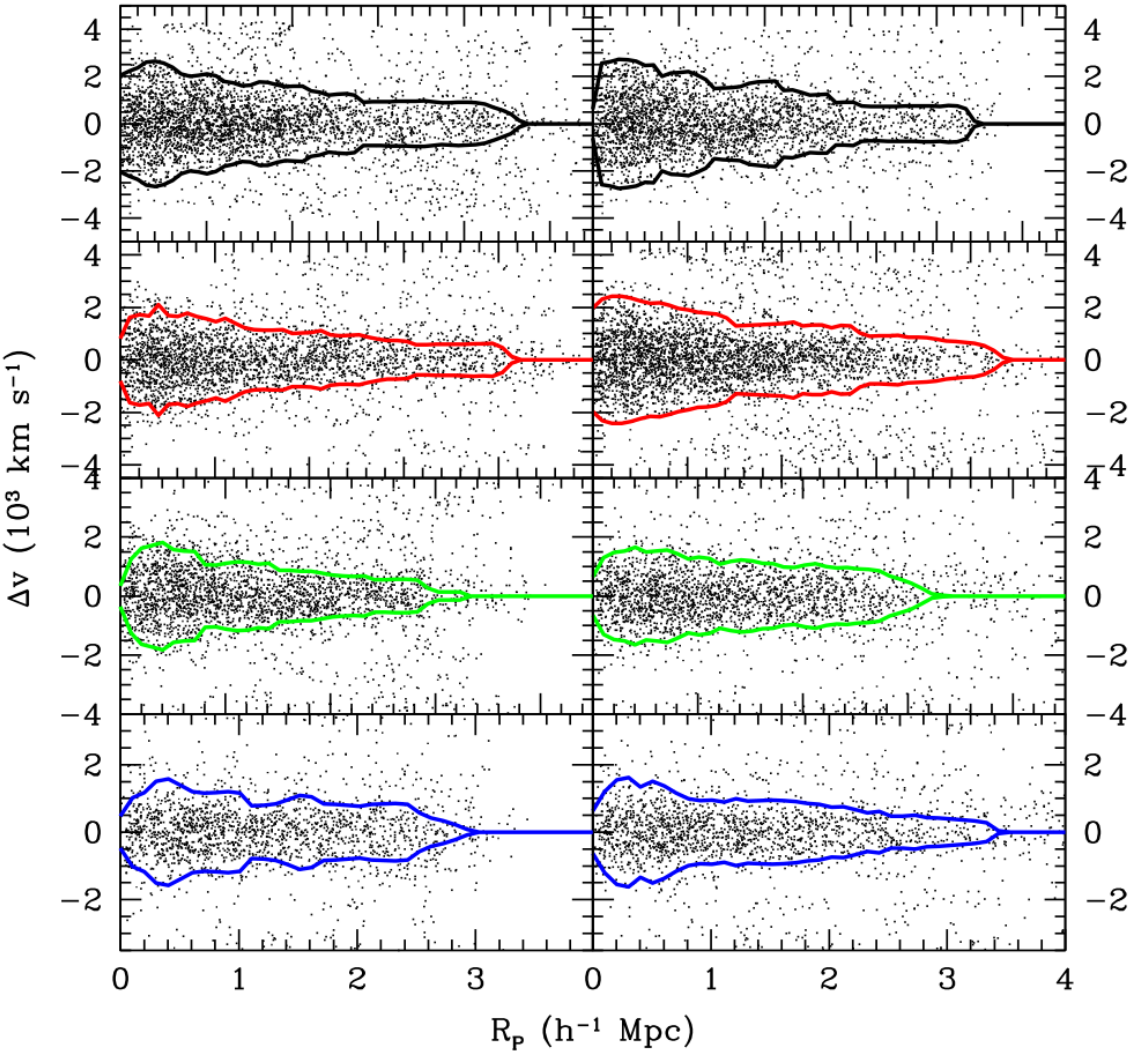


Figure 1: Example of caustics in ensemble clusters, where multiple clusters have been stacked together according to their X-ray luminosity quartiles on the left and M_{200} quartiles on the right side. X-axis has projected radius and y-axis has the velocity offsets from the cluster center. The solid colored lines indicate the caustic mass profiles. Galaxies inside the area are very probable cluster member galaxies and outside most likely not tied to the cluster gravitational potential. Luminosity/mass is decreasing from the top to the bottom if the figures. [13]

2.3.6 Sunyaev-Zel'dovich

The Sunyaev-Zel'dovich (S-Z) effect was named after two astrophysicist from the Soviet Union who studied the effect in papers released in 1969 and 1972. The S-Z effect is the inverse Compton scattering of the cosmic microwave background (CMB) by ICM. When CMB photons pass through massive galaxy clusters, they encounter very energetic electrons in the gas of the cluster. They scatter with the electrons and gain energy from the encounters, which shows in the CMB background as slightly

hotter areas. Since the distortion of the CMB spectrum traces the temperature of the ICM, and the temperature correlates with mass, the S-Z effect can be used to detect galaxy clusters and to estimate their masses. S-Z observations need to be backed up by distance measurements from optical observations. Photometric observations may be sufficient, but getting spectra of the clusters will give more precise redshifts. [14]

2.4 Spectroscopic properties of galaxy clusters

Optical emission in galaxy clusters comes mainly from the galaxies in them, and the optical emission in galaxies comes mainly from the stars in the galaxies, so we have to know what the spectra of stars are like in order to know about the spectra of galaxy clusters. In this thesis the spectroscopic interest is in elliptical galaxies, since they are the main component in clusters besides the inter galactic medium (IGM). I have chosen an example from the SPIDERS data of a elliptical galaxy residing in a cluster at redshift $z = 0.44$ (SPIDERS ID: 1_10157), which can be seen in figure 4 and a heavily smoothed and zoomed in portion of it in figure 5. In the figures, restframe wavelengths are shown at the top and observed wavelengths at the bottom. The spectrum is showing many emission and absorption features as well an underlying continuum emission. The details of the spectrum are explained in the next subsections.

2.4.1 Cluster galaxy continuum emission

The main feature of a normal, old, red main sequence star residing in an elliptical galaxy, is a black body continuum spectrum determined by the surface temperature of the star. Figure 2 shows a template of a typical G type star that can be expected to be found in elliptical galaxies. The spectrum is a black body, peaking roughly around 5000 Å. The spectra for an elliptical galaxy is an integration of all it's gas and stars, which at intermediate redshifts are typically some Gyr old red main sequence stars - similar to the sun - and red giants. There can be some star formation going, but a lot of the gas in galaxies is stripped by the cluster gas when galaxies plunge to the central parts of clusters, which will effectively prevent star burst events in galaxies. In SPIDERS, the galaxies are targeted from red sequence galaxies (see section 3.3.2), so the spectra are by default mainly those of red galaxies. [8]

2.4.2 Emission lines

Emission lines from stars are weak. The strongest emission lines come from very hot stars like O and B stars. Since these are blue giants, they have hardly any contribution to the emission from elliptical galaxies. Weak emission lines from red stars are buried in the noise if we are observing further than our solar neighbourhood. Merely looking at the G type star in figure 2, it is clear that it is not showing any

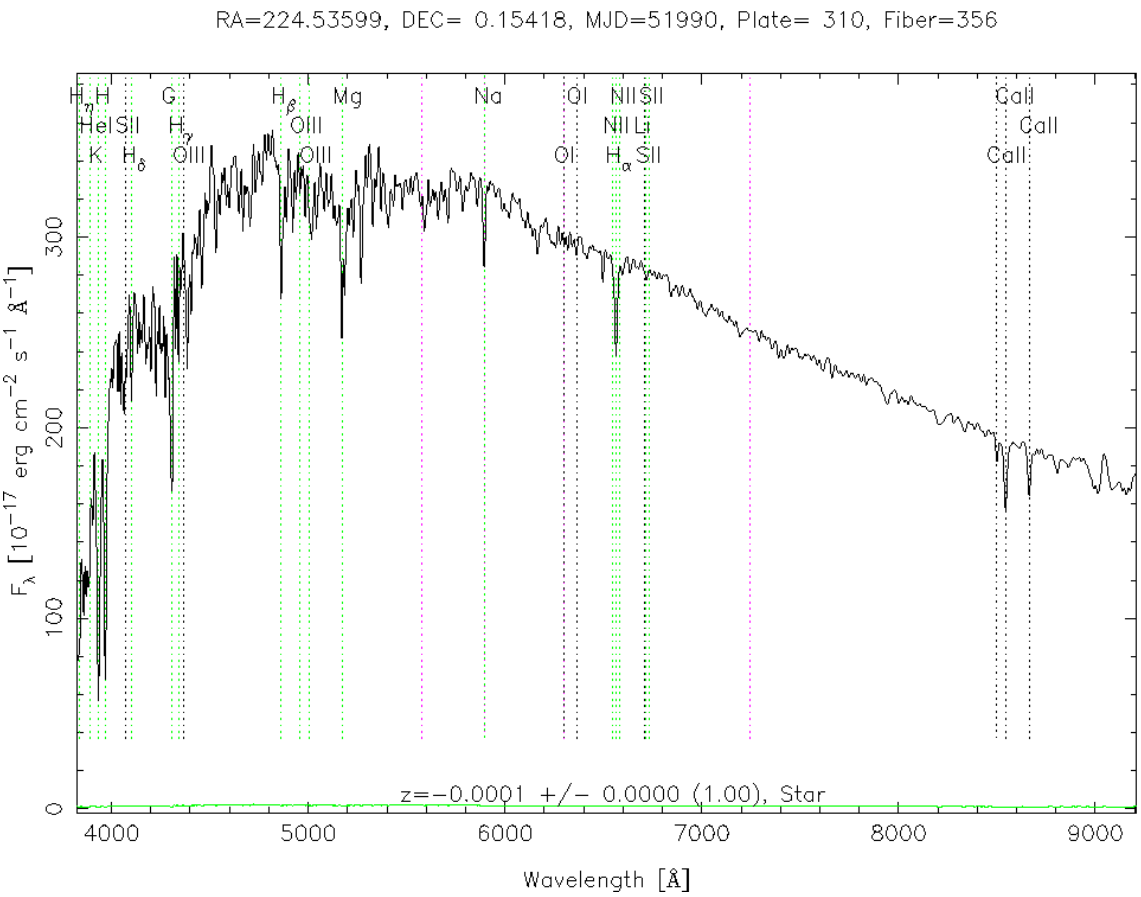


Figure 2: The spectrum of a typical G type star is a black body continuum peaking around 5000 Å with absorption lines. G type stars and other red and old stars dominate the spectrum at intermediate redshift elliptical galaxies. [15]

emission lines, which would stand out, so we don't expect to see them at redshifts $z > 0.1$ either.

There are many emission lines in observations though. Many of these come from the Earth's atmosphere and they are usually called sky emission, airglow or sky spectra. Airglow at night time happens when energy gathered from the Sun is released high in the atmosphere (80 to 110km high). When UV photons arrive from the Sun, they hit particles in the atmosphere causing atoms to excite or ionize and molecules to dissociate. When the Sun sets, the atoms and molecules start going back to their ground states and recombining, and in the process releasing the excess energy as photons. Common sources of emission lines are for example recombining O_2, relaxation of OH vibrational states, many sodium molecule reaction, etc. Figure 3 shows the median airglow of the BOSS spectograph taken from data gathered in the years from 2009 to 2017. (Read more about BOSS in section 3.2.1.) Since the atmosphere is in constant motion due to winds, density variation and other local conditions, sky emission has changes at different time scales from seconds to years. Also differences in the Sun's activity affect sky emission lines. [16]

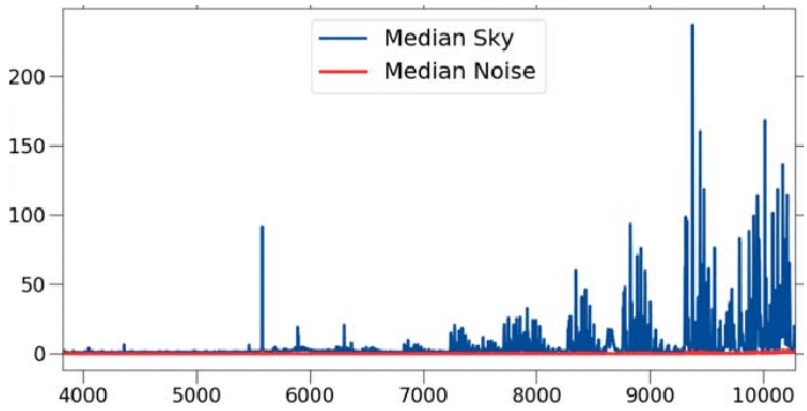


Figure 3: BOSS median airglow taken in the years from 2009 to 2017. [16]

In figure 4, strong sky emission lines are clearly visible at both blue and red ends of the spectrum. Both figures 4 and 5 are also showing the strong atomic oxygen transition at 5577 Å and the sodium D doublet at 5889 and 5895 Å. [17]

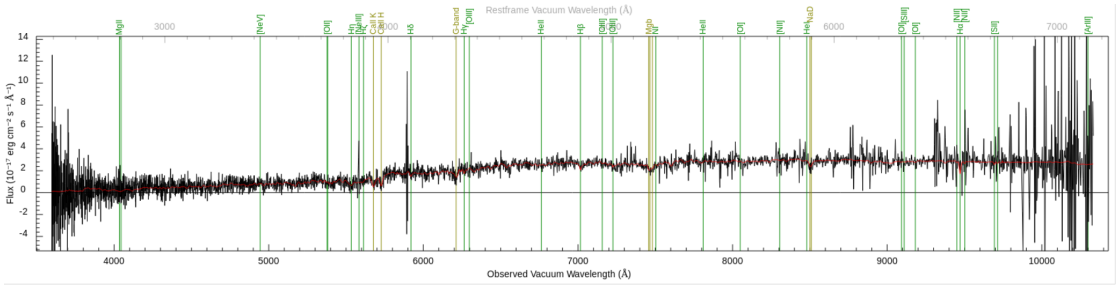


Figure 4: Spectrum of a galaxy at redshift $z = 0.44$. The black line indicates observed spectrum and the red line is the model spectrum of a suitable galaxy. At the top is the resframe wavelengths and at the bottom the observed ones. The green lines mark up the usual spectral lines found in galaxies. The main feature is a underlying continuum, with a lot of absorption and emission lines. The absorption comes from stellar atmospheres, ISM, IGM, and the Earth's atmosphere. Most of the emission lines are sky emission or other noise which are especially significant at both the red and blue ends of the observed spectrum.

2.4.3 Absorption features

Absorption of photons from stellar origin happens in different mediums from stellar atmospheres to intergalactic gas and dust, and all the way to the Earth's atmosphere. This section describes the main features of absorption from different origins.

The main identifying absorption features to look for in red galaxies are the 4000 Å break and Ca II H and K lines. Less important, but still possibly visible features are the Balmer break and many specific absorption lines, that I will go through next.

The 4000 Å break is a drop in intensity at restframe wavelength $\lambda < 4000$ Å. It is

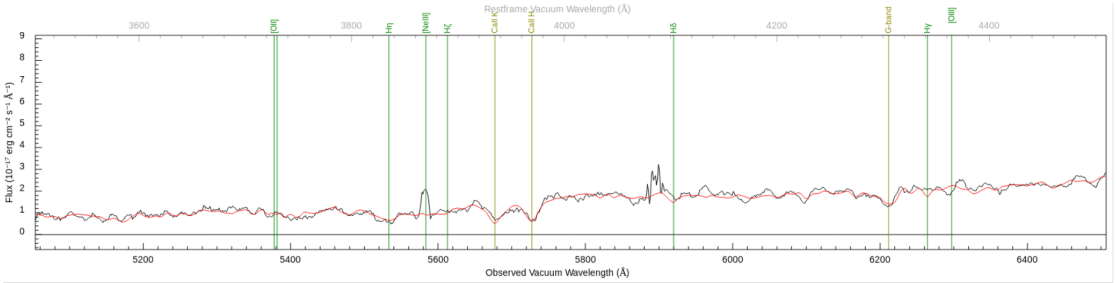


Figure 5: Heavily smoothed spectrum of a galaxy at redshift $z = 0.44$. The image has been zoomed in to show especially the Ca II K and Ca II H lines and the 4000 Å break. Details are the same as in figure 4.

created by absorption of high order Balmer series lines and many different metals at stellar atmospheres. The strongest lines are calcium lines of Ca II K and H at 3969 Å and 3934 Å. The effect is intensified in old star populations, where metallicities are higher. Also noteworthy is that the ratio of Ca II K and H lines can tell about the age and metallicity of the stellar population. When the Ca II K line is stronger than Ca II H, then the old stellar population is dominating the spectrum over younger stars and vice versa.

The Balmer break is produced by multiple absorption lines of the Balmer series at wavelengths $\lambda < 3650$ Å. It is present in most galaxies, but is less visible in old stellar populations like in elliptical galaxies. Figure 6 has an example of two galaxy spectra, where both the 4000 Å and Balmer breaks are visible. The black solid line shows the spectra of a very red galaxy with an old metal rich stellar population and the dotted line shows a much bluer galaxy with younger stars, including O and B giants. The Balmer break is very clear in the bluer galaxy, but it is hard to detect in the red one. The 4000 Å break with Ca II K and H lines on the other hand is the most visible feature in the red galaxy, where it is not so noticeable in the bluer one.

G-band absorption at 4295 to 4315 Å can be seen clearly in figure 4. The name G band is a relic from the turn of the 20th century and comes from a German astronomer Fraunhofer who studied the Sun's spectrum and named the absorption lines with letters. The feature is a combination of many spectral lines, most of which belong to the CH molecule in stellar atmospheres. [19] Sun like stars are plentiful in elliptical galaxies so it's logical to expect G band absorption to be found in their spectra as well.

Mg_b lines from 5154 to 5197 Å are present in metal rich stellar populations. The amplitude depends also somewhat on the age of the stars. Hβ from 4848 to 4877 Å is the opposite to Mg_b, because the line depends mostly on the age and in a lesser capacity on the metallicity of the stars. Both lines are visible in the example galaxy spectrum in figure 4

Na I D absorption line at 5891 Å to 5896 Å is produced by the interstellar medium (ISM). The same medium is responsible for overall reddening of galaxy spectra. Tiny dust particles and gas are hit with photons, which heats up the medium and photons

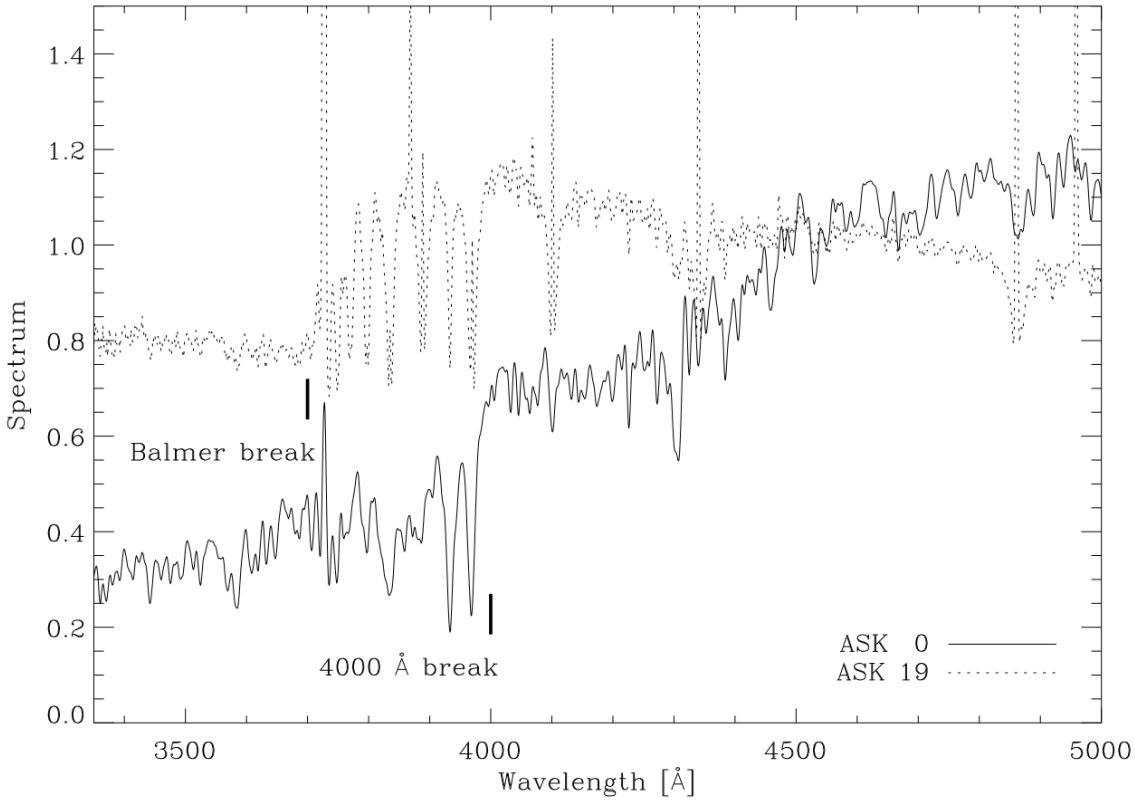


Figure 6: Balmer and the 4000 Å breaks in two different galaxies. In black solid is a typical red elliptical and in the black dotted line is a much bluer galaxy. [18]

scatter with lower energy than the original. This is called extinction and it happens in all galaxies. Variations in extinction depend on how much dust is between the object and observer. [18][8]

2.4.4 Redshift related spectral properties and observational limitations

Besides extinction from dust which logically increases with increasing space between the object and observer, there are also other effects from distance, or more specifically from the expansion of the volume of space between. Redshift happens when wavelengths lose energy in the expansion as $1 + z = \lambda_{observed}/\lambda_{emitted}$, where z is redshift and $\lambda_{observed}$ and $\lambda_{emitted}$ are the observed and emitted wavelengths.[8] All electromagnetic radiation moves to longer wavelengths, meaning the light from distant clusters becomes redder. For example, the 4000 Å break moves into 10000 Å at redshift $z = 1.5$. Infrared observations can still be done from specific places on Earth, at high altitude deserts and mountain ranges, but it is difficult due to the opacity of the atmosphere. Around 10000 Å, it becomes very challenging to observe with ground based optical telescopes. Figure 7 shows the quantum efficiencies for filter bandpasses of the SDSS Apache point telescope, which is where the optical data for this thesis comes from. In the figure it is easy to see that 10000 Å is at the edge of the possible wavelength range.

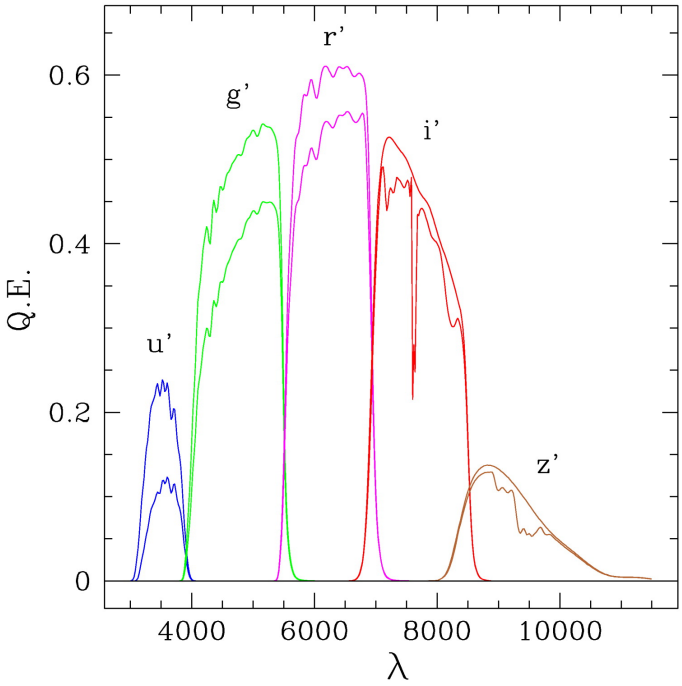


Figure 7: SDSS filters. The upper curves correspond to optical throughput of the filters and the lower curves to expected atmospheric extinction with airmass of 1.2. QE on the y-axis stands for quantum efficiency, meaning the sensitivity of the CCD. [20]

The other aspect of limitations due to distance to the object, is that apparent brightness F of objects drop according to $F = L/(4\pi d^2)$, where L is luminosity and d is the distance to the object. [8] Objects become too dim to be observed with reasonable integration times and with Earth based telescopes. This apparent brightness limit happens at lower redshifts than the redshift limit due to reddening described earlier in this section, so the apparent brightness limit is the main distance related limitation for observing galaxy clusters. The limits of optical ground based observations in SDSS-IV and SPIDERS for clusters without major activity like AGN are around $z = 0.7$ and even those will be the most massive clusters. This can be seen from figure 8, which represents the CODEX cluster sample i-band magnitudes at different redshifts (see section 3.4 for details of CODEX). Also shown in the figure is that the majority of clusters are at i-band magnitudes under 20 and the limits are roughly at magnitude 21.

The concentration of galaxies within the virial radius r_{200} has no significant trend in redshift. The luminosity function does not change much with redshift either. It also has a similar shape for different masses of clusters, but the smaller clusters of course have less galaxies in them compared to very massive clusters, which logically decreases the chance of observing them. [21][22]

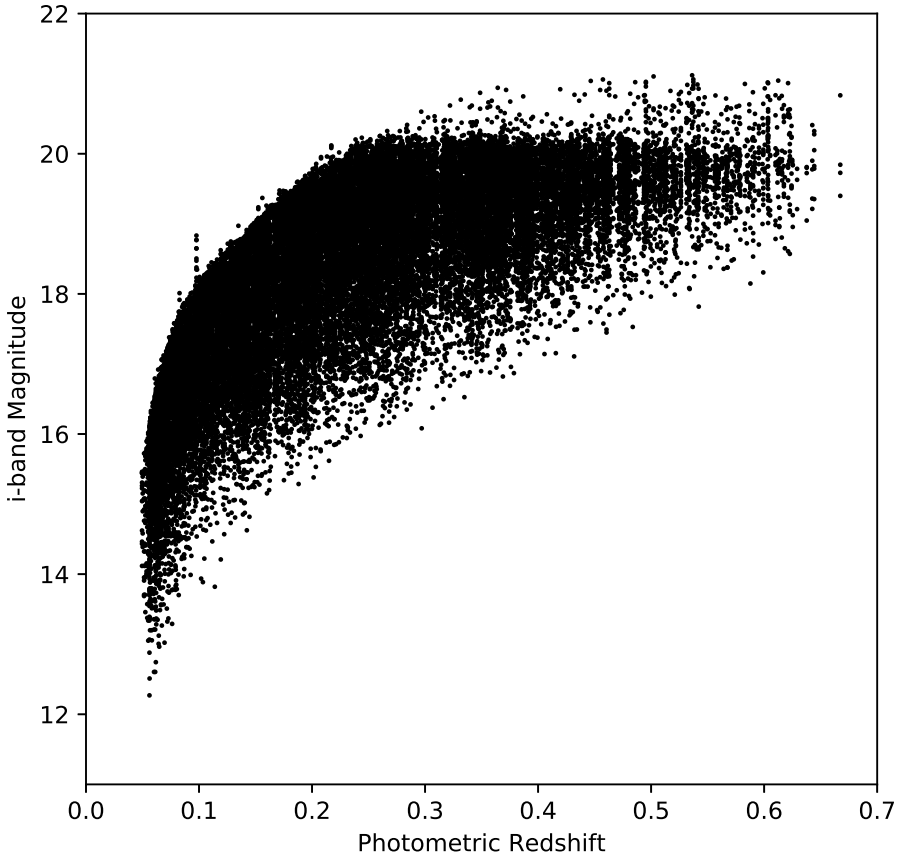


Figure 8: The i-band photometric magnitudes versus photometric redshift in the CODEX cluster sample. The limiting i-band magnitude for these observations is roughly 21 with the majority of clusters detected at brighter than 20 mag.

3 Data and methods

3.1 SPIDERS program

SPIDERS stands for SPectroscopic IDentification of eROSITA Sources. SPIDERS is an observational program with the primary goal of doing spectroscopic follow up of extragalactic X-ray selected sources. SPIDERS obtains spectra of both extended sources and point-like sources. A target list for SPIDERS clusters comes from the CODEX catalogue photometric observations (see section 3.4 for CODEX). [10] The point-like sources are mainly active galactic nuclei (AGN) which will not be covered in this thesis.

SPIDERS was planned for following up of sources detected with the X-ray telescope eROSITA. The X-ray sources for SPIDERS come from an old X-ray survey: the

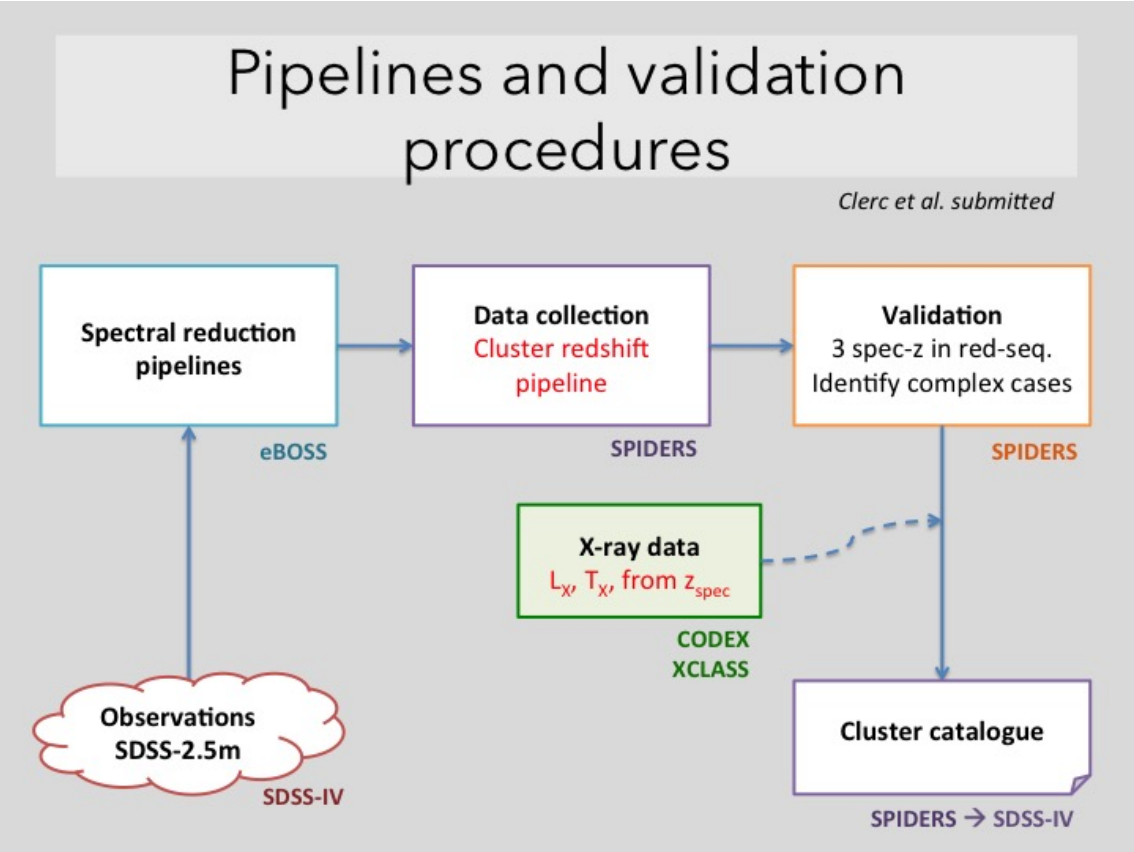


Figure 9: Pipelines and validation procedures for making the SPIDERS value added catalogue

ROSAT all-sky survey. These were the likely first pass X-ray detections of eROSITA. As the launch of eROSITA was delayed many times from the original launch date in 2014 to 2019, the sources have not been available for SPIDERS catalog as was the plan and ROSAT remains the source of galaxy cluster X-ray detections. More about the X-ray data used in this thesis will be covered in section 3.4. Details of the optical data can be found in section 3.2. Observations for identifying optical counter parts for the X-ray sources are done with the Sloan Digital sky survey (SDSS) imaging data. As a short description, an algorithm called the redMaPPer (described in section 3.3.2) was run to identify over-densities of galaxies at the same photometric redshifts, creating the CODEX catalogue optical counterparts. These were matched to the X-ray sources and then probable member galaxies were assigned to be observed by SDSS with the BOSS spectrograph mounted on their telescope at Apache Point observatory. The main survey is called eBOSS which stands for extended Baryonic Oscillation Spectroscopic Survey, where SPIDERS is a sub-program (see section 3.2.1 for eBOSS). eBOSS conducted reductions on the spectroscopic data through their own pipeline. Figure 9 provides an overview of the different pipelines and procedures from observations to the final SPIDERS cluster catalogue.

The SPIDERS program produced a value added catalogue which was released to the public as a part of SDSS-IV final data release. Compared to photometric cluster finding methods like redMaPPer, spectroscopic observations can disentangle line-of-sight projection effects based on the galaxy redshifts. Details of the automated process are given in section 4.1. This automated process is not enough however and the data needs to be validated. Validations are done by people with the help of an online tool. The validation process is explained extensively in section 4.2. The final redshift estimate for the cluster is given by the bi-weight average of member galaxy redshifts. See section 3.6.1 for the details on the bi-weight method. Statistical uncertainties Δ_z are estimated with a bootstrap re-sampling of redshifts. The uncertainties are typically $\Delta_z/(1+z) \lesssim 10^{-3}$ for a median number of member galaxies $N_{mem} = 10$. [10]

3.2 Optical data: Sloan digital sky survey (SDSS)

The Sloan digital sky survey is one of the most extensive surveys in astronomy. SDSS routine operations started in 2000 and are still continuing. SDSS uses a 2.5 meter telescope situated at the Apache point observatory in New Mexico, USA. Instruments used at the telescope are a 250 megapixel CCD camera and a fiber-optic spectrograph. Imaging is done in multiple optical bandpasses: ultraviolet (u), green (g), red (r), near-infrared (i) and infrared (z). (See also figure 7.) So far SDSS imaging covers a third of the sky. Image data is free to the public.

Spectral imaging started in 2009 with the Baryonic oscillation spectroscopic survey (BOSS) instrument. Current generation of SDSS is SDSS-IV (2014-2020), which includes among others the eBOSS program in which SPIDERS is a sub-program.

3.2.1 extended Baryon Oscillation Spectroscopic Survey

Within SDSS there are different programs. One of these is the extended Baryon Oscillation Spectroscopic Survey (eBOSS). The survey's main science goal is to map out the epoch when the Universe switched from gravitationally bound deceleration to the current accelerating expansion of the Universe. Main science questions for eBOSS are related to constraining dark energy, growth of structure, evolution of quasars at high redshifts and constraining neutrino mass.

eBOSS observations concentrate on red galaxies and quasars up to redshift $z = 3$. The observations for SPIDERS clusters lay in the redshift range from $z = 0.01$ to $z = 0.7$. The galaxies in clusters are not all high luminosity galaxies, which is why they start disappearing from view at higher redshifts and also start to shift from optical wavelengths into the infrared range as redshift increases, the further we look into history. Luminous red galaxies and emission line galaxies may be observed up to $z = 1.0$ and $z = 1.1$ respectively and quasars in the range $0.8 < z < 3.5$. There are no previous large spectroscopic surveys in this redshift range so eBOSS fills a gap and will be part of making a more complete data set in conjunction with other

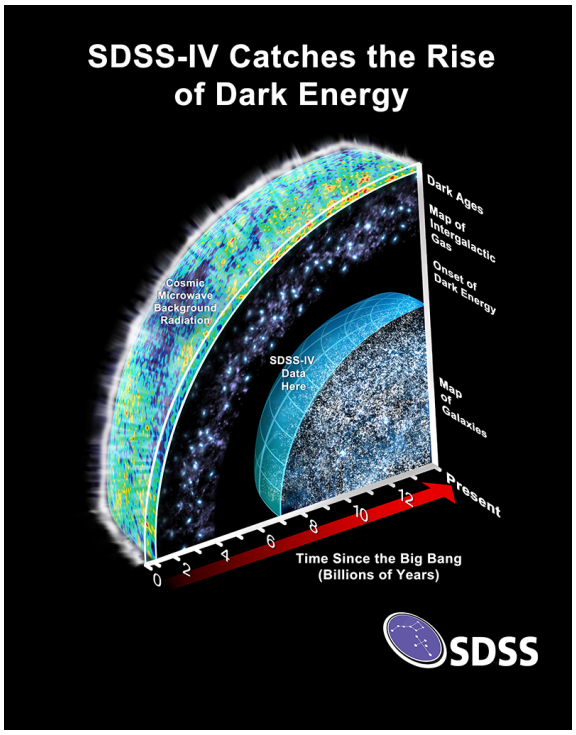


Figure 10: eBOSS program. Image credit: Dana Berry / SkyWorks Digital Inc. and the SDSS collaboration.

surveys. [23]

Observations have been made in 2014-2020 at a wavelength range of 3600-10000 Å. The BOSS instrument used for eBOSS uses custom made fiber-optic plates. Each observation plate has 1000 fibers placed manually into holes drilled in specific design patterns according to where the targets are in the sky. The fiber allocations are divided between SPIDERS clusters and AGN, The Time-Domain Spectroscopic Survey (TDSS) for variable objects and main eBOSS targets. Fibers can only be at a certain distance from each other. The aperture of a fiber for eBOSS is 2" and minimum distance between fibers is 62". Some areas of the sky are gone over more than once, so there can be multiple passes over single SPIDERS clusters. The targeting scheme for SPIDERS is explained in section 3.5 The redshifts of the galaxies for SPIDERS are obtained from the automated eBOSS pipeline which identifies targets from the shape of their spectra. It is quite straightforward, since the galaxies all fall into the same red, passive galaxy types. The spectra show clear absorption features like the 4000 Å break, Ca II K and H, G-band and Na D lines, which are identified and redshift is calculated from their wavelengths. See section 2.4.3 for details of the absorption features. [10][23][24]

3.3 Finding clusters in data

Galaxy cluster finding algorithms in general search for some form of overdensity of mass. Whether it's counting photons (X-ray) or the change in CMB temperature (SZ), counting optical targets or examining the bending of light (weak lensing) depends on the method. In this section I will introduce two specific methods that were crucial for making of the CODEX catalogue, which is the base for the SPIDERS cluster catalogue targets.

X-ray sources are found by searching for photon overdensities. Many kinds of sources show X-ray emission in varying scales and energies. Galaxy clusters are not point sources, but spatially extended if the resolution and sensitivity of the instrument is sufficient. To find galaxy clusters from the data, a good option is to map out the overdensities in different scales, and the ones with wider profiles than the point-spread-function of the instrument, can be called extended. A specific technique for finding galaxy clusters, called the wavelet detection algorithm, is introduced in the following section 3.3.1. [5]

Photometric galaxy cluster finding methods all count galaxies. This basic idea can be achieved by running simple algorithms. The main issue with photometric overdensity counts is that they may count random projection effects. For example, many small galaxy groups who just happen to be in the line of sight, appear as a rich galaxy cluster. To counter this major problem, estimating redshifts is usually the answer. Some methods use estimated redshifts for each galaxy, some work in color space, comparing different filter bands to estimate redshifts of the galaxies or clusters. It's not straightforward though, because galaxy clusters are not all the same, some have starburst galaxies or spiral galaxies and they may be experiencing in-fall or mergers. These affect the colors of galaxies and morphologies of galaxy clusters and since photometric redshifts are derived from the difference of emission in different filter bands, having very varying colors in galaxies will affect the results. However, it has been found that almost all galaxy clusters share one type of galaxies: old, red, early type elliptical galaxies. To find these red ellipticals, only two filter bands of observations need to be taken as long as the trademark of early type ellipticals - the 4000 Å break - is visible. This technique is called the red sequence method. [25] The one used for finding CODEX clusters is developed from this basic idea and it's called redMaPPer. Details of the algorithm are given in section 3.3.2.

It's also possible to do blind spectroscopy in search of galaxy clusters, but the efficiency of such methods may be questioned since spectroscopy takes more observation time than photometry and the chance of observing foreground or background objects is quite obvious.

3.3.1 Wavelet detection algorithm

Wavelet detection algorithm was created to find X-ray sources from data. The idea of the wavelet detection is to have a convolution with a kernel that consists of a positive core and an outer negative ring, with the integral of the kernel over the x,

y plane being zero. A general form for a kernel is given in equation 1, where r is the radius and the scale $a = 1/2b$ is varied to the wanted values. An example of how the wavelet detection works is shown in figure 1 where the kernel is convoluted with a Gaussian $s(r) = exp(-r^2/2\sigma^2)$. In figure 11(a), is an image with two objects of different scale overlapping each other. A simple detection algorithm works well to find a single object, but with multiple objects in the same image only the higher peak would be discovered like in figure 11(b) even with multiple scales. With a more advanced wavelet detection, objects are found as local maxima in the convolved images and an iteration of detections are reduced from the original image. This cleaning process is continued on all the wanted scales and at the end both sources are modelled instead of only the stronger point source, see figure 11(c). In 11(d) the smallest three and largest three scales are summed and the final model achieved. In addition to finding sources, the algorithm also gives the scale of the sources, which can be useful for determining the type of the source. AGN are point sources and clusters are more extended sources when the observations have sufficient resolution for determination. [5]

$$W(r) = \frac{exp(-r^2/2a^2)}{2\pi a^2} - \frac{-r^2/2b^2)}{2\pi b^2} \tag{1}$$

3.3.2 redMaPPer algorithm

Rykoff et al. have developed a complex iterative method for finding clusters of red sequence galaxies in a large amount of photometric data. They have named it Red-sequence matched-filter probabilistic percolation cluster finding algorithm (redMaPPer). RedMaPPer works with photometric colors, using two or more filter bands of data in search of the most prominent feature in early type elliptical galaxies, the 4000 Å break, where the flux coming from the galaxies changes significantly. It performs efficiently with self-training from an initial small spectroscopic training set. It can run on very large data sets with a broad redshift range across several filters, and gives membership probabilities to all galaxies. It also has a richness estimator that is designed to minimize scatter, which is needed for cosmological studies. First, I will describe the richness estimation and after that the cluster finder.

Cluster membership probabilities are calculated with:

$$p_{mem} = p(\boldsymbol{x}) = \frac{\lambda u(\boldsymbol{x}|\lambda)}{\lambda u(\boldsymbol{x}|\lambda) + b(\boldsymbol{x})} \tag{2}$$

where $\boldsymbol{x}$ is a vector of observable properties explained later in the text, λ is the number of galaxies in the cluster (ie. the richness of the cluster, see section 2.3.3), $u(\boldsymbol{x}|\lambda)$ is a density profile of the cluster, which is normalized to unity, and $b(\boldsymbol{x})$ is the background and foreground density of galaxies not belonging to the cluster.

The richness λ of the cluster is defined as the sum of the membership probabilities:

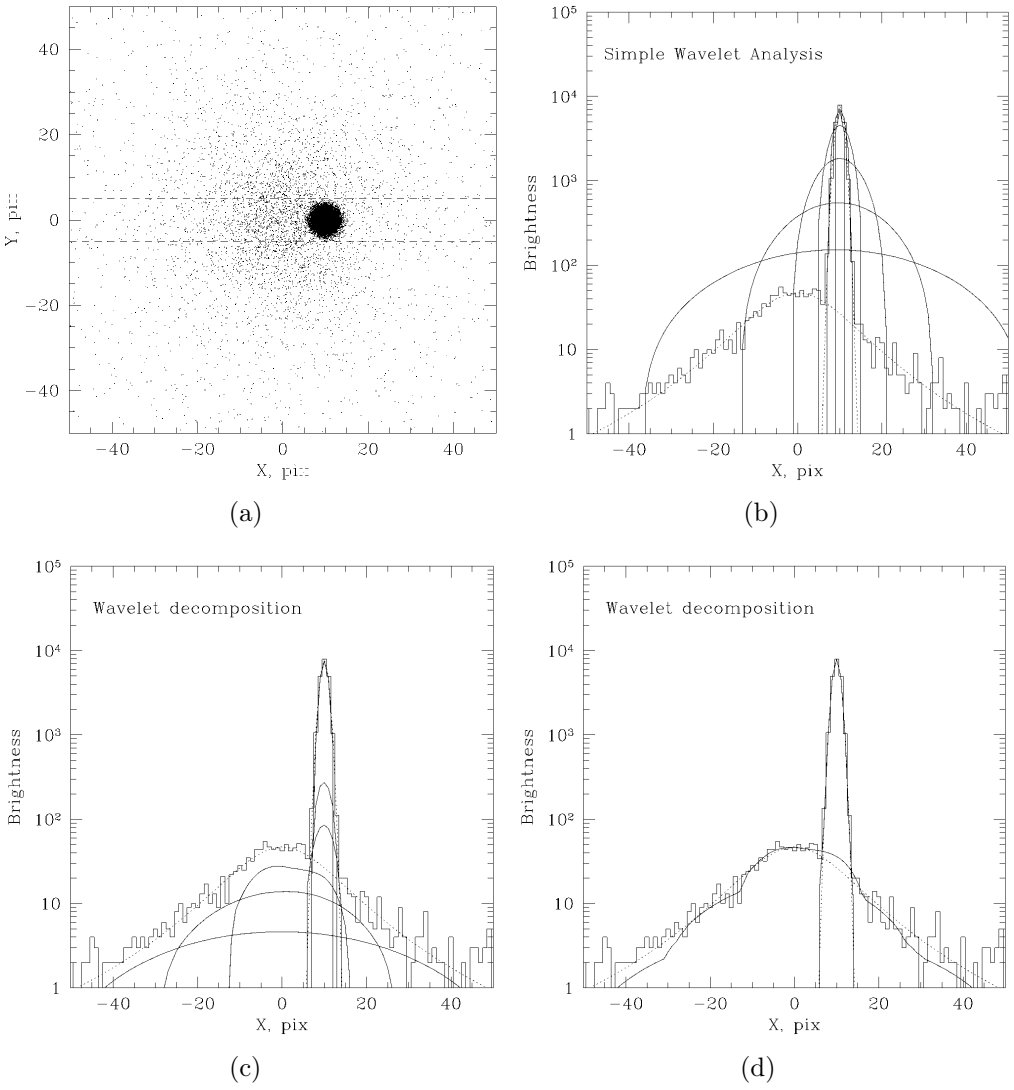


Figure 11: Example of how wavelet detection works on simulated data. 11(a) shows the simulated data. 11(b) is how a simple wavelet detection has only found the strong point source in a multiple component image. Using a more complete wavelet algorithm, like in 11(c), is advantageous as it shows how an iterative process can clean the image, finding both sources in the data. In 11(d), the three smallest and the three largest scales are summed together to model both sources well. [5]

$$\lambda = \sum p(\boldsymbol{x}|\lambda) = \sum_{R<R_c(\lambda)} \frac{\lambda u(\boldsymbol{x}|\lambda)}{\lambda u(\boldsymbol{x}|\lambda) + b(\boldsymbol{x})} \tag{3}$$

where R_c is the cluster radius estimate. R_c is not any standard overdensity radius like R_{500c}, but a optical cutoff radius for the redMaPPer algorithm. It scales with richness as follows:

$$R_c(\lambda) = R_0(\lambda/100.0)^{\beta} \tag{4}$$

where $R_0 = 1.0h^{-1}$ Mpc and $\beta = 0.2$. Richness is solved numerically by minimizing equation 3. The statistical uncertainty of λ is given by:

$$\mathrm{Var}(\lambda) = \sum p(\boldsymbol{x}|\lambda)[1 - p(\boldsymbol{x}|\lambda] \tag{5}$$

The $\boldsymbol{x}$ in the previous equations contains three observational properties of the galaxy in question: the radial position, the luminosity and the color. These three properties are included into the vector using filter functions: two dimensional cluster galaxy density profile $\Sigma(R)$, the cluster luminosity function $\phi(m)$ and the χ^2 distribution with ν degrees of freedom $\rho_\nu(\chi^2)$, where ν is the number of colors.

The radial filter $\Sigma(R)$ scales as:

$$\sum(R) \propto \frac{1}{(R/R_s)^2 - 1} f(R/R_s) \tag{6}$$

where the characteristic scale radius is $R_s = 0.15h^{-1}$ Mpc and $f(x) = 1-(2/\sqrt{x^2-1})$ $\tan^{-1}\sqrt{(x-1)/(x+1)}$. The core density is assumed constant and the radius is truncated at the cluster radius R_c.

The cluster luminosity function $\phi(m)$ follows a Schecter function that depends on the magnitude as:

$$\phi(m_i) \propto 10^{-0.4(m_i - m_*)(\alpha+1)} \exp(-10^{-0.4(m_i - m_*)}) \tag{7}$$

where $\alpha = 1.0$ and $m_*(z)$ is a polynomial that has two forms depending on which side of redshift $z = 0.5$ the object is:

$$m_*(z) = \begin{cases} 22.44 + 3.36\ln(z) + 0.273\ln(z)^2 \\ -0.0618\ln(z)^3 - 0.0227\ln(z)^4, & \text{if } z \leq 0.5 \\ 22.94 + 3.08\ln(z) - 11.22\ln(z)^2 \\ -27.11\ln(z)^3 - 18.02\ln(z)^4, & \text{if } z > 0.5. \end{cases} \tag{8}$$

The luminosity function is normalized to unity at cutoff luminosity and characteristic magnitude: $0.2\mathrm{L}_*$, and m_* 1.75 mag.

$\rho_\nu(\chi^2)$ describes the difference of the galaxy model color to the observation, given the i-band magnitude and redshift. The usefulness of the χ^2 distribution is that it is a single variable instead of the full color vector, which may have 5 or more filter bands. Using a single variable makes the computation less time consuming. In the following equation 9, $\mathbf{c}$ is the galaxy color vector, $\mathbf{c}|z, m_i)$ is the mean of the red sequence model color at the wanted redshift z and i-band magnitude m_i. The scatter and correlations of the galaxy colors about the mean are taken into account in the covariance matrix $\mathbf{C}_{int}$ and the photometric error of the galaxy is described in the error matrix $\mathbf{C}_{err}$.

$$\chi^2(z) = (\mathbf{c} - (\mathbf{c}|z, m_i))(\mathbf{C}_{int}(z) + \mathbf{C}_{err})^{-1}(\mathbf{c} - (\mathbf{c}|z, m_i)) \tag{9}$$

The red sequence cluster members then have the distribution:

$$\rho_\nu(\chi^2) = \frac{(\chi^2)^{\nu/2-1}e^{-\chi^2/2}}{2^{\nu/2}\Gamma(\nu/2)} \tag{10}$$

where ν is the number of colors used in the color vector.

The three filters function are joined into a single filter function through:

$$u(\boldsymbol{x}) = [2\pi R\Sigma(R)]\phi(m_i)\rho_\nu(\chi^2) \tag{11}$$

where $2\pi R$ is needed to get the probability density distribution of $\Sigma(R)$.

The background, meaning all non cluster member galaxies, is estimated as a uniform density: $b(\boldsymbol{x}|z) = 2\pi R\bar{\Sigma}_g(m_i, \chi^2|z)$ where $\bar{\Sigma}_g(m_i, \chi^2|z)$ is the galaxy density given the i-band magnitude and χ^2 distribution at the appropriate red sequence model redshift. The mean galaxy density is estimated by making a grid of model galaxies, which are brighter than $0.1L_*$, in χ^2 and i-band magnitude space with a redshift spacing of 0.02. Then calculating χ^2 for all galaxies of the catalogue and binning them with a well-known grouping algorithm called cloud-in-cells. The binned galaxies are then divided by the survey area and placed into the grip. Then finally the galaxy number density is normalized by the width of each color and magnitude bin.

RedMapper also accounts for magnitude and geometrical limitations of clusters and the occasional bright stars or other artefacts as masked areas. The estimate of number of masked galaxies is done by Monte Carlo sampling of an appropriate model galaxy cluster and getting the mean of galaxies which are found in the masked area. This is then applied to the richness estimate of the observed red sequence cluster.

I describe next the cluster finder phase of RedMaPPer. The algorithm starts with an initial set of spectroscopic "seed galaxies", which are taken from prior observations (for example SDSS Data release 8). These galaxies are often the bright central galaxies of clusters. First, redMaPPer searches for similar color galaxy overdensities around these seeds in a single color. With SDSS data the appropriate filter band in redshifts below 0.35 is the g-r band and for redshift above 0.35 the r-i band.

As a first estimate of a red sequence, the spectroscopic seed galaxies are placed in redshift bins of width 0.025, then estimating with a Gaussian fit the mean and the intrinsic width. Galaxies within 2σ range from the estimated mean are considered red. The binned galaxies are used to find the mean color-redshift relation. The means are then used to find the width of the color-redshift relation and finally a clean sample of seed galaxies can be chosen within 2σ from the calculated mean of the model color at the spectroscopic redshifts. This process gives a full color-redshift model from spectroscopic redshifts as seen in figure 12. The black dots are the seed galaxies, the red dashed line is the model fit, the red dots are the mean colors at redshift intervals of 0.1 and the blue dash-dotted line indicates redshift $z = 0.35$ where the divider for the 4000 Å break is located in the SDSS filter colors. It's easy to see the scatter increases considerably on both of the "wrong" sides of the filter 4000 Å break line.

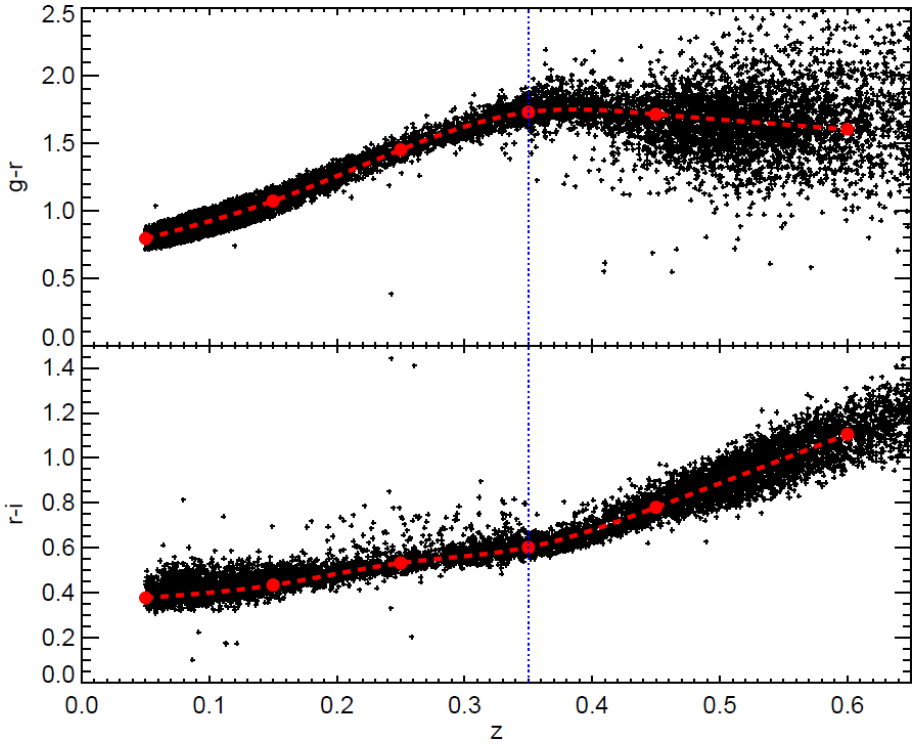


Figure 12: Example of a redMaPPer red sequence model from spectroscopic seed galaxies. The black dots are the spectroscopic seed galaxies, the dashed red line indicates the found model and the red dots are points of median color placed every redshift 0.1 from the start of the sample. The blue dash-dotted line indicates the redshift point $z = 0.35$ where the different SDSS filter colors (g-r or r-i) can be used to sample the 4000 Å break. [26]

In the next phase of the algorithm, single color red sequence memberships are determined with the help of the seed galaxy color-redshift model. A color box with a width of 2σ and radial range of 500 h^{-1} kpc is selected around each spectroscopic seed galaxy. The σ width for galaxies with $z \leq 0.35$ (SDSS color g-r) is 0.05 and for $z \geq 0.35$ (SDSS color r-i) it is 0.03. Red sequence is then fitted to these galaxies

and a single color richness is estimated. Membership probabilities are calculated with equation 2 for overdensities found with a richness $\lambda_{singlecolor} > 10$. This single color is achieved by using a slightly different full color filter with ν =1 instead of the χ^2 described in equation 10. Photometric galaxies with a membership probability $p_{mem} > 0.7$ (see equation 2) are given the seed galaxy redshift at this stage. This calibration is important to sample also the lower magnitude galaxies, since the spectroscopic seed galaxies tend to be the brightest of their clusters. This stage produces a training sample with an accurate amplitude, tilt and scatter of the red sequence for the next calibration phase.

The next phase produces a final model with all available filter colors at the full range of redshifts and magnitudes. To make sure only red galaxies are sampled instead of blue, a color cut is made at a 1.5σ range, where σ is the median absolute deviation from the color median. To avoid biases on scatter with the exclusion on blue galaxies, the cut is implemented also in the likelihood function. Galaxies with membership probabilities $p_{mem} > 0.7$ are taken and a new fitting procedure is done. To get a result, an iterative process must be done, as the membership probabilities depend on the red sequence parameters (amplitude, tilt and scatter) and the red sequence parameters depend on membership probabilities. The iteration converges quickly and only after 3 iterations the calculated richness is affected only by $\sigma < 0.1$. At this point all member galaxies are still associated with the spectroscopic seed galaxies.

To go from red sequence clusters determined using seed galaxies to estimating properties of actual photometric clusters, an accurate redshift determination is needed. To do this, an initial redshift of the central galaxy needs to be estimated. The log-likelihood of the color distribution for a cluster is:

$$\ln\mathcal{L} = -\frac{\chi^2}{2} + \ln \mid \frac{dV}{dz} \mid \tag{12}$$

where χ^2 is defined in equation 9 and the term $\frac{dV}{dz}$ is included because the volume of space is larger at higher redshifts. The likelihood is maximized along a redshift grid to find the best central galaxy candidate redshift value. This is applied as the initial redshift guess for the cluster redshift.

For the cluster redshift, another likelihood function is determined as:

$$\ln\mathcal{L} = \sum -\frac{w[\chi^2 + \ln|\mathbf{C}|]}{2} \tag{13}$$

where $\mathbf{C}$ is the covariance matrix of galaxy color properties and w is the weight of each member galaxy such that w is between 1 and 0:

$$w(p_{mem}) = \frac{1}{\exp[(p_{70} - p_{mem})/0.04] + 1} \tag{14}$$

where p_{70} means the threshold probability which accounts for 70% of the total

richness and the 0.04 is where the redshift distribution width is set. Using weights is preferred instead of a hard cut to overcome instabilities in the iteration and a 70% is chosen so that the same fraction of galaxies is used regardless of redshift or richness. Using a hard p_{mem} cut would lead to less galaxies making the cut at higher redshift, because of the noisier photometric data leading to lower membership probabilities.

Richness and membership probabilities are calculated at each step. High membership probability galaxies are chosen (p_{70}) and new redshift value is estimated with the maximizing the likelihood of equation 13. This iterative process is continued until it converges to less than $\Delta z < 0.0002$.

Centering a cluster is important for an accurate richness estimate and especially for cosmological implementations with the resulting catalogue. Miscentering is a known issue for redMaPPer, but there isn't a better algorithm developed for it so far. Miscentering occurs mostly if the central galaxy (CG) of the cluster is having massive star formation, which makes it too blue to be considered for red sequence, or the center is chosen as a red foreground galaxy.

Centering is done as an iterative process like redshift estimation. The initial guess for a CG is simply the brightest galaxy with a membership probability of at least 0.8. Each galaxy is estimated to achieve a probability P_{cen} and P_{sat} to be a central galaxy or a satellite galaxy. The filter functions for centering are a magnitude and richness dependent luminosity filter (ϕ_{cen}), a Gaussian photometric redshift filter ($G_{cen}(z_{initial})$) and a local galaxy density filter ($f_{cen}(w)$), which gives weights to galaxies depending on their radius from the center, luminosities and membership probabilities. Satellite and foreground galaxies get their own similar functions. In the same initial iteration, the filter functions get a first estimation. The subsequent iterations are done by maximizing a likelihood function that is based on filter functions for the three different galaxy types possible: the central galaxy, satellite galaxy and the foreground/background galaxy.

The overall process of cluster finding is pictured in the flow chart in figure 13. In the first pass, possible galaxy cluster centers are mapped, first with generous cuts in $\chi^2(z_{initial}) < 20$ and L $> 2.0L_*$, and then demanded that at least 3 suitably red galaxies are found for each overdensity. Cluster initial redshifts and richnesses are calculated for the candidates. Likelihood sorting takes place next. Likelihoods are calculated with:

$$\ln\mathcal{L} = \ln\mathcal{L}_\lambda + \ln\mathcal{L}_{cen} \tag{15}$$

where the centering likelihood is:

$$\ln\mathcal{L}_{cen} = \ln[\phi_{cen}(m_i|z_\lambda,\lambda)G_{cen}(z_{initial})f_{cen}(w|z_\lambda,\lambda)] \tag{16}$$

and the richness likelihood is:

$$\ln\mathcal{L}_\lambda = -\frac{\lambda}{S} - \sum \ln(1 - p_{mem}) \tag{17}$$

where S is a scale factor, which is a correction term for estimated fraction of galaxies masked out at the current redshift.

The amplitude of the richness likelihood is usually much larger than the centering likelihood. When two candidate central galaxies have similar richness likelihoods then the centering likelihood becomes meaningful in the outcome of the most likely center.

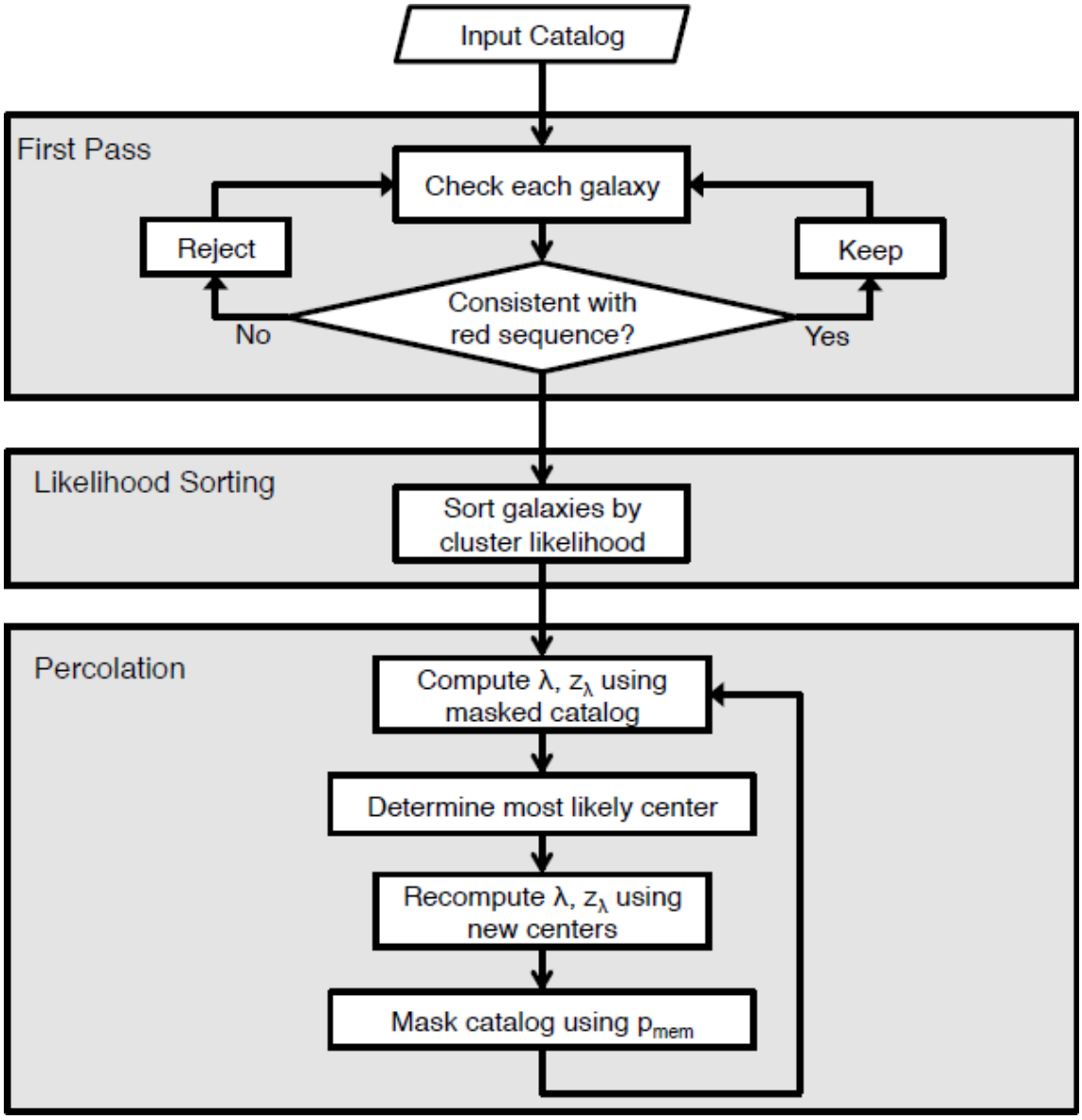


Figure 13: Flow chart for the redMaPPer cluster finder. [26]

Next phase is the percolation of the likelihood ordered galaxy cluster candidates. Candidate clusters are taken as input and calculated for richness and redshift. Cen-

tering is done and a new estimate is calculated for richness and redshift according to the new most likely center. Member galaxy lists are updated according to their membership probabilities and unlikely centers with $p_{mem} > 0.5$ are removed from the current cluster (but they can still be part of other candidate clusters).

The result of the finished algorithm is a red sequence selected galaxy cluster catalogue, with richnesses, redshifts and errors. [27][26]

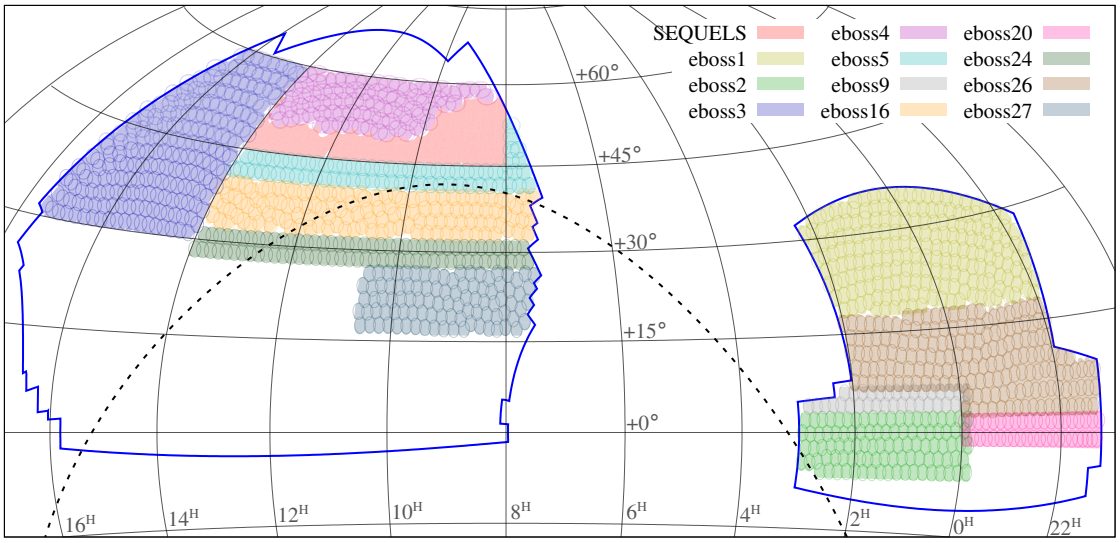


Figure 14: This figure shows the eBOSS chunks which have at least one target common with SPIDERS. Each circle is representing an observation plate of 3 degrees, with 1000 optical fibers. The blue lines are the boundaries of the BOSS optical imaging area and the black dashed line is the divider between German and Russian halves of the sky of eROSITA data. SEQUELS is part of the SPIDERS catalogue, but it was observed already earlier in SDSS-III and has more overlap than the rest of the chunks.

3.4 CODEX catalogue: link between X-ray and optical data

The CODEX (COnstraining Dark Energy with X-ray) program is a survey where faint X-ray sources from the ROSAT all-sky survey (RASS) were selected and the redMaPPer algorithm (described extensively in section 3.3.2) was run with SDSS photometric imaging data to search for optical counterparts for the objects.

The ROSAT x-ray satellite flew 30 years ago and was the first satellite to perform an all-sky survey at 0.1-2.4 keV energy band. Main observations were done from 1990 to 1991 and covered most of the sky. Some missing patches were observed later in 1997 and ROSAT was turned off in 1999. [28].

RASS data release 3 is the basis for the X-ray data used in making the CODEX catalogue. It has 41 square degree count rate maps and background count rate

maps, which have some overlap, for several different energy bands. An example of a background count rate map is in figure 15. Average depths for the whole sky are 400 seconds. Lowest fluxes possible for RASS, reach 10^{-13} erg s^{-1} cm^{-2}. Only a few photons are needed for a reliable detection due to extreme sensitivity and low background of the telescope. [28][29]

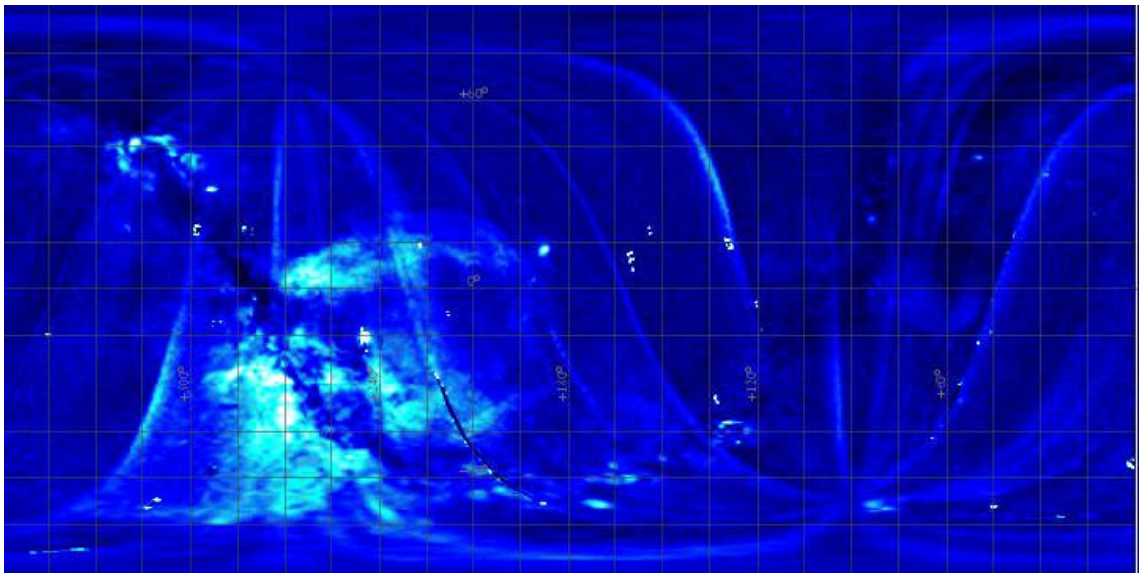


Figure 15: RASS full-sky background count rate map for energy band 0.4 - 2.4 keV.

Figure 16 shows the X-ray sensitivity within the SPIDERS footprint in equatorial coordinates. Sensitivity is slightly better in the northern sky as indicated with blue color in the figure. Some areas have worse sensitivity colored in gray and red, but these are fairly small patches and the overall area can be considered uniform. [30]

For finding CODEX sources from RASS data, the wavelet detection algorithm described in section 3.3.1 was used. Threshold for detection was set to 4σ, which is close to the lowest possible threshold. Detections can be as low as 4 photon counts.

The search for X-ray sources was done in different scales from 1.5' to 12'. Then the small scales and large scales detection catalogues were merged and duplicates removed. Sources with offsets below 3' are considered duplicates. The reason for using different scales is to make sure all possible sources are found like in the example described in the wavelet detection algorithm section 3.3.1. Using larger scales than 12' may result in false sources from background variations, and are not used. Estimation of cleanness of the catalogue is 0.998. In a million sources, there were 53 false detections calculated from probability. 90% are AGN and 10% are galaxy clusters.

Optical counterparts were identified at the source areas with the redMaPPer photometric red-sequence code, described in section 3.3.2. The highest richness areas were chosen within the redshift range $0.05 > z > 0.8$. Best optical center - BCG of the cluster - was identified within 400 kpc of the X-ray center. Richnesses were calculated in two positions: one at the X-ray center and one at the optical center.

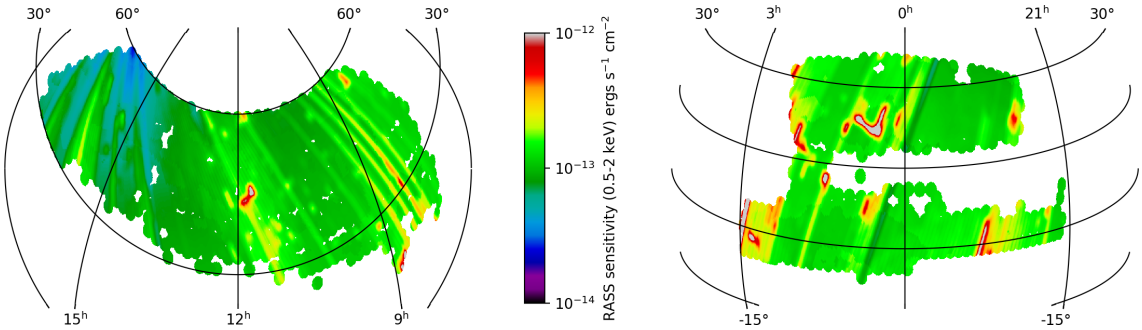


Figure 16: RASS sensitivity within the SPIDERS footprint. The survey area is 5350 square degrees. On the left is the norther galactic cap and on the right, the southern galactic cap. The round dots are the SDSS fiber plates each one 3 degrees in diameter. There is a higher sensitivity in the northern galactic cap area marked with blue and some local low sensitivity regions in gray where the data was mostly unusable, but in general the area is very uniform. [30]

CODEX catalogue is concentrated on the faint sources in RASS data, because all large studies done before it were focused on the bright sources. For photometric identification, 10 members were chosen as the minimum amount of galaxies for consistency, as other redMaPPer catalogues have used the same in the past. As one can guess, this limit results in a rising richness with redshift since the likelihood of observing smaller galaxies drops with increasing redshift (as explained in section 2.4.4). Figure 17 shows the photometric sensitivity of the CODEX survey. It is quite uniform in the whole 5350 square degree area. The area is the same in both figures 16 and 17. The round dots are the SDSS fiber plates which are 3 degrees in diameter. [30]

Doing scientific work with the CODEX catalogue is already possible: With a redshift to assign to the X-ray sources, one can estimate local X-ray properties of the clusters, like X-ray luminosity and mass of the clusters. For SPIDERS, the contribution of the CODEX catalogue was to serve as a starting point to assign targets for spectroscopic observations through membership probabilities.

3.5 Targeting scheme for SPIDERS

X-ray sources were selected from the RASS faint catalogue that were within the full eBOSS footprint area. The goal of SPIDERS was to observe all the counterparts with optical spectroscopy which had not yet been observed. Optical targets to be observed using spectroscopy within the SDSS-IV were chosen from the CODEX catalogue described earlier. The highest probability members were assigned to have a fiber in the SDSS-IV fiber plates. There were 1137 fiber plates, each with 3 degree diameter, observed in the duration of the program. Each plate had 1000 fibers connected to the BOSS spectrograph. Only a small fraction of the fibers were allocated to SPIDERS and the majority to eBOSS targets, but since SPIDERS targets are much more rare on the sky than eBOSS targets, they were given high

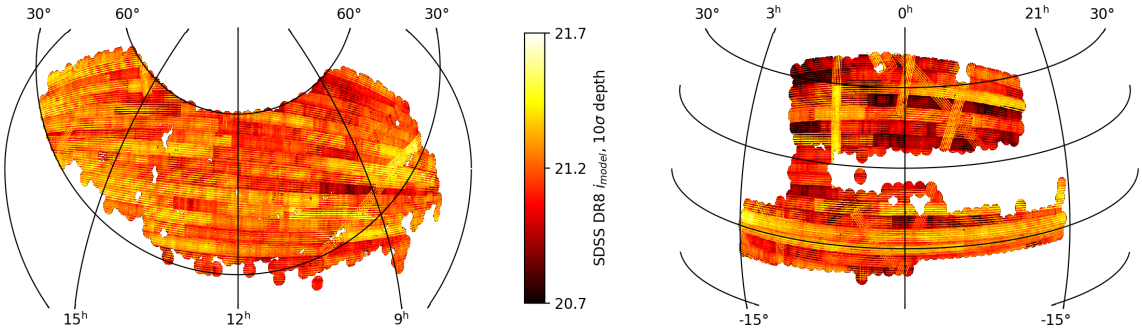


Figure 17: SDSS photometric i-band sensitivity within the SPIDERS footprint. The survey area is the same 5350 square degrees as for the X-ray data in figure 16. Similarly on the left is the norther galactic cap and on the right, the southern galactic cap. The round dots are the SDSS fiber plates each one 3 degrees in diameter. The survey photometric sensitivity is quite uniform although the limiting magnitude is lower in the southern sky. [30]

priority. There is overlap in the observations so dense target areas could have more fibers than the minimum distance of 62" between fibers would otherwise allow. Typical galaxy clusters in the redshift range of the SPIDERS survey range a few arcminutes in the sky. Red sequence galaxies tend to concentrate at the centers of clusters according to their luminosities, so it's important to have overlapping plates to get an adequate number of luminous galaxy spectra for each cluster. [10][30]

SPIDERS spectra was done quite uniformly within the area. Two chunks had slightly different targeting strategy. For chunk eboss3 the fiber allocation was denser regarding SPIDERS targets to accommodate the higher sensitivity in RASS data for that area. And for chunk eboss20, seen in figure 18 as pink circles, the targeting had extended cluster centered radii to up to five times the virial radius for a few selected targets. This was done so that part of the data could be used for caustic method mass determination. See more about caustics in section 2.3.4. Figure 19 shows the X-ray flux curves plotted against sensitivity. Except for chunk eboss3, they are almost uniform. [30]

3.6 Velocity dispersion methods

3.6.1 Bi-weight

The bi-weight variance estimator has been available for astrophysicists for a few decades. It has been tested extensively and has been found to be a robust (meaning the estimate stays about the same when the distribution from which the data points are drawn is varied) and resistant (meaning the estimate stays about the same when a number of data points are replaced by other values) estimator for sample variances. The bi-weight variance is:

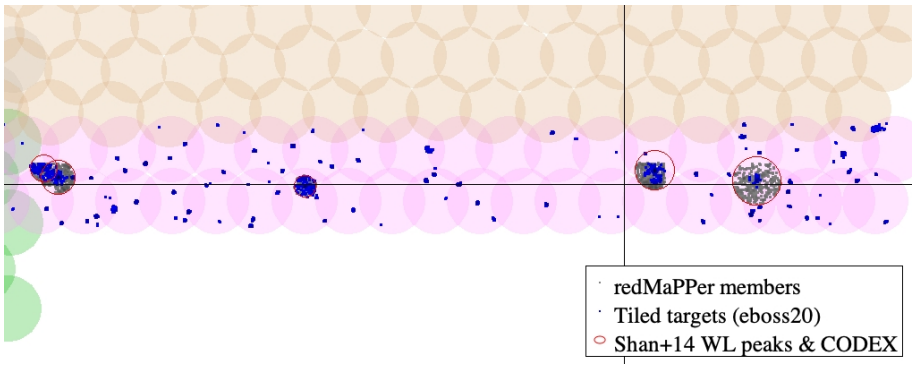


Figure 18: A map of redMaPPer red sequence members in gray and spectroscopic eBOSS targets in blue in the chunk eboss20. This part of the sky is part of stripe82, which was observed more than other areas of the survey, so it has more targets for SPIDERS. The image is a zoom of figure 14 at declination 0 degrees and right ascension 23h. The red circles indicate 5 times the virial radius of the clusters in CODEX and CFHT weak lensing maps. [30]

$$\sigma_{BI}^2 = N_{members} \frac{\sum_{|u_i|<1}(1-u_i^2)^4(v_i-\bar{v})^2}{D(D-1)} \tag{18}$$

where v_i are the proper velocities, $\bar{v}$ is their average and D is:

$$D = \sum_{|u_i|<1}(1-u_i^2)(1-5u_i^2) \tag{19}$$

where u_i is the bi-weight weighting:

$$u_i = \frac{v_i - \bar{v}}{9\mathrm{MAD}(v_i)} \tag{20}$$

where MAD is an acronym for median absolute deviation, MAD = median($|v_i - \bar{v}|$). Using median instead of mean makes the formula more resistant to interlopers, because if there are any, they are found at the minimum and maximum values and do not affect the median like they do the mean. 9 is a tuning constant that seems a good value for balance of efficiency in velocity dispersion calculations according to previous studies.

Bi-weight scale estimator works well for sample sizes of 15 members or more. It starts failing at lower member numbers if the distribution has any significant deviation from a Gaussian distribution. It is the recommended scale estimator for galaxy clusters with at least 15 members and is used in SPIDERS and this thesis for those. [11][31]

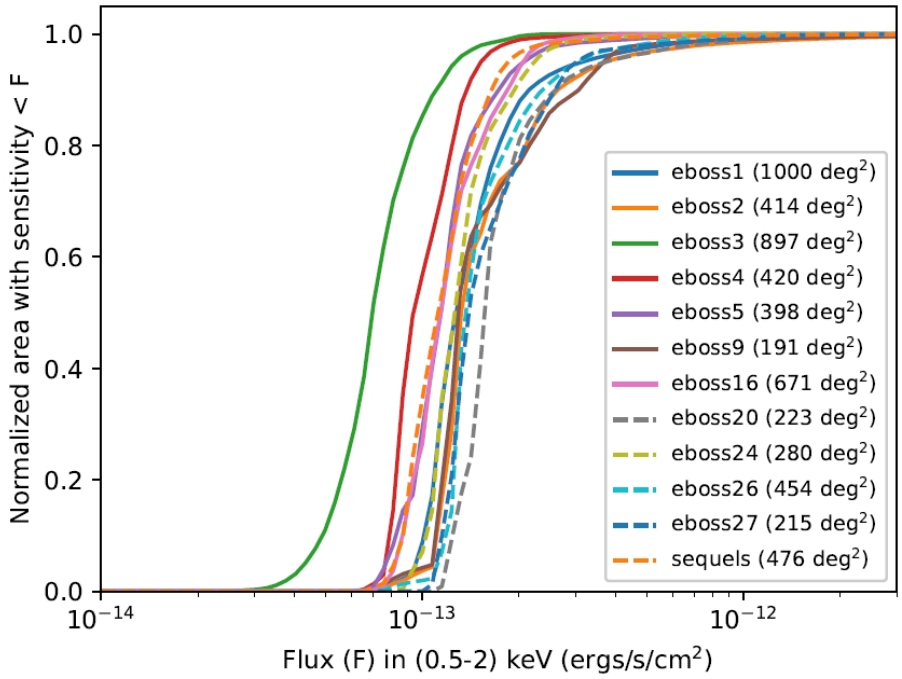


Figure 19: X-ray flux in the 0.5-2.0 keV band in the different chunks of SPIDERS survey area plotted against normalized sensitivity. Except for better sensitivity in chunk eboss3, they are quite uniform. [30]

3.6.2 Gapper

The gapper scale estimator is a different kind of estimator compared to many other ones. It is based on the gaps in an ordered sample, where weights are given to the data points according to their location on the ordered list.

Gaps are defined as:

$$g_i = v_{i+1} - v_i, i = 1, ..., n-1 \tag{21}$$

and weights as:

$$w_i = i(n-i) \tag{22}$$

The gapper scale estimator is defined as:

$$\sigma_G = \frac{\sqrt{\pi}}{n(n-1)} \sum_{i=1}^{n-1} w_i g_i \tag{23}$$

The gapper method is resistant to interlopers and performs better than other methods in small samples if the distribution is non-Gaussian. Detailed simulations and studies have been conducted in the past, which recommend the gapper method for

sample sizes of lower than 15. Although for samples of $n = 5$ or smaller the values may be questioned, but this is true for all methods of scale estimation. In general, scatter increases when the sample size decreases. [11][31]

4 Analyzing data

4.1 Automated spectroscopic cluster finding process

To alleviate manpower from a very tedious process of going through all observations by hand, an automated process is used before the validation of the objects in the cluster catalogue. The SPIDERS sample contains clusters with a range of redshifts from 0.01 to 0.7, which have a variety of different masses and velocity dispersions. To accommodate this, a broad selection is used to pick up cluster candidates from the observations.

The process starts from a bi-weight average of all members of the red-sequence galaxies, which have a spectroscopic redshift. All candidates with a velocity offset of more than 5000 km/s are rejected. This limit is higher than studies usually have, which is to make sure a larger number of the cluster galaxies are included. Then an iterative process is run with a 3σ clipping of the average velocity dispersion until it converges or 10 iterations are done. Velocity dispersion method depends on the number of members. Bi-weight variance is used for members $\geqslant 15$ and Gapper velocity dispersion for members < 15. See sections 3.6.1 and 3.6.2 for details of the methods. [10]

4.2 Spectroscopic cluster validation

The automated pipeline for cluster member selection and cluster redshift determination works well. Most objects can be validated as they come out from the pipeline. More complex cases need an educated person to inspect what qualifies as a validated cluster, how multiple components are handled, which galaxies are included in the cluster and what is considered contamination. To ease the validation process the SPIDERS team has developed a online web tool which is called the screening web interface.

The screening web interface has a multitude of options to help the user make educated decisions. Figure 20 has an example of plots that the web interface has to offer for a well sampled cluster. Plots available directly in the screening interface are the following; 20(a) a ”membership probability” plot with spectroscopic redshift reduced from the redMaPPer cluster redshift, 20(c) a ”projected distance to centre - velocity offset to central redshift” plot, two sky position plots 20(b) and 20(d): one with all available spectra and one with currently selected members, 20(e) a spectroscopic redshift histogram and 20(f) a ”velocity offset to central redshift” histogram. The membership probability plot shows all galaxies in the specific plate and their cal-

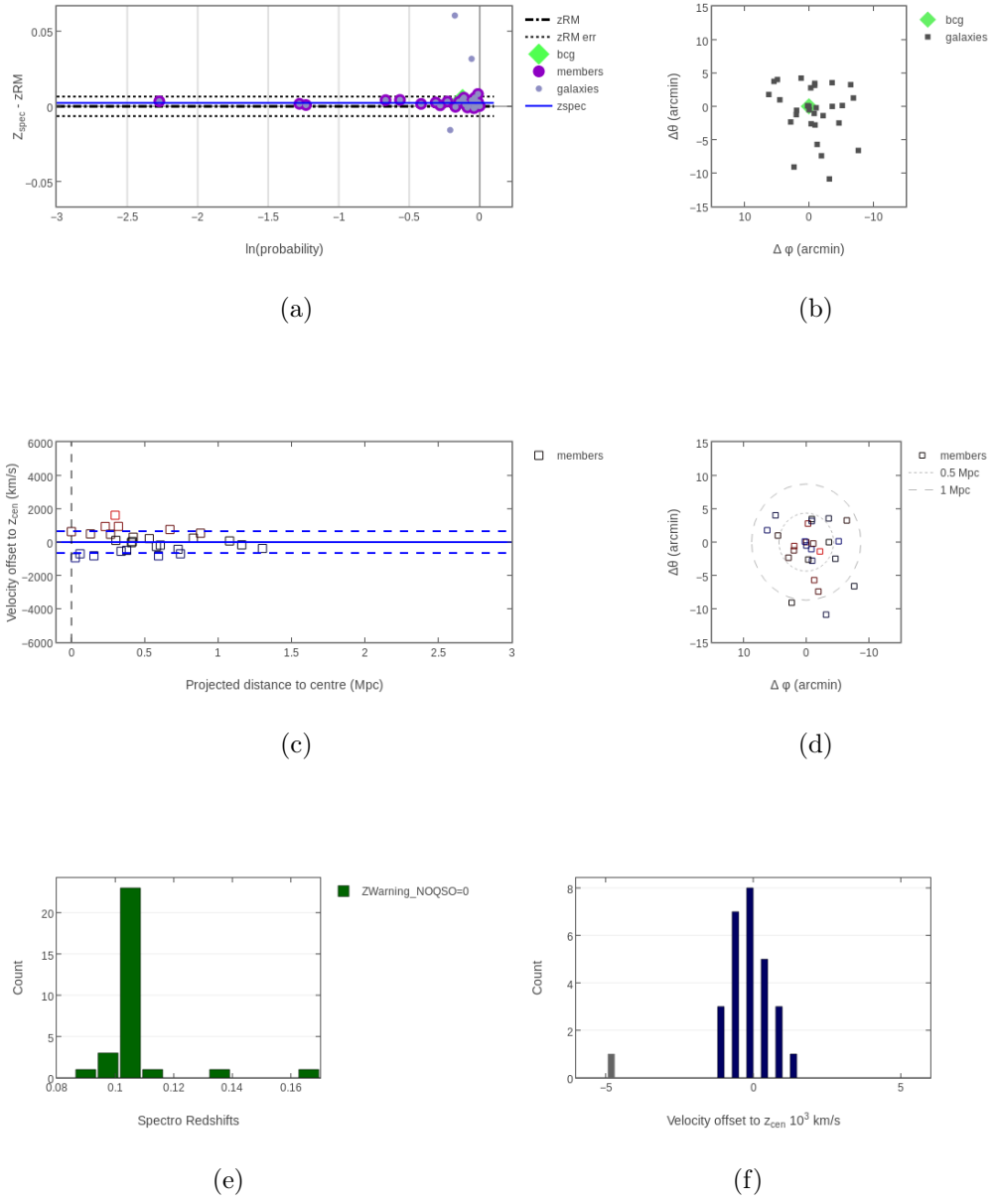


Figure 20: The SPIDERS screening web interface plots for galaxy cluster 1_10183: (a) logarithmic membership probability plot with spectroscopic redshift reduced from the redMaPPer redshift; (b) ”projected distance to centre - velocity offset to central redshift” plot; (c) sky-plot with all available spectra; (d) sky-plot with currently selected members with color-coding; (e) spectroscopic redshift histogram; and, (f) ”velocity offset to central redshift” histogram. The colors in plots 20(c) and 20(d) indicate velocity offsets to the normalized zero point which in this case means the bi-weight average of the member velocities. Green in plot 20(e) means galaxies and blue is for QSO’s. Blue in histogram 20(f) means the currently selected members and gray the non-selected ones. Circles in the sky-plots indicate the estimated distances from the center position.

culated probability to be a part of the cluster on the x-axis. It highlights members selected for the cluster and highlights specifically the suspected BCG of the cluster. Bad galaxy spectra are flagged. On the y-axis is spectroscopic redshift with the redMaPPer redshift as zero. redMaPPer redshift errors are also shown in the plot. In the "velocity offset - projected distance" plot, only selected members are plotted. X-axis has the projected distance to the chosen center in units of Mpc and the y-axis has a velocity offset to the central redshift in units of km/s. Galaxies are colored with regards to their velocity offsets from the bi-weight average (marked with a blue solid line), which helps in visualizing the target. Blue dashed lines are the velocity dispersion values ($\pm 1\sigma$). There is a sky positions plot with all spectra from the plate and highlighting the possible BCG. Origin on the plot is set on the chosen center and-axis units are in arc minutes with zero at the center. The second sky position plot is compiled in the same manner as the first one, but only with selected member galaxies, which are colored with the same color coding as plot 20(c). To help the user, there are also concentric circles to indicate estimated distances from the center. The spectroscopic redshift histogram shows the number counts from the plate under investigation. Galaxies are colored with green and quasi stellar objects (QSO's) are colored with blue. The velocity offset histogram shows number counts of galaxies in each bin. Units are in 10^3 km/s with the zero point being the bi-weight average of the member galaxy redshifts.

Users can check information about the target in tabs directly on the screening interface (see figure 21). Information menu in figure 21(a) includes richness at the optical centre (BCG), richness at the X-ray centre, redMaPPer and spectroscopic redshift of the cluster (bi-weight average), optical and x-ray center coordinates, observation status in the next menu (figure 21(b)) with information of what survey the galaxy spectra are from. The third menu tab in figure 21(c) shows the currently selected member galaxies and maybe most importantly the velocity dispersion and spectroscopic redshift calculated with the current selection. Velocity dispersions are calculated with two methods: The Gapper method and the bi-weight method. The Gapper method is used when there are less than 15 members and the bi-weight method is used for calculation of at least 15 members. In the last top menu tab (figure 21(d)) is shown the current velocity clipping parameters that can be re-calculated in the side menus which are described below.

Additional information can be accessed through menu options. In figure 22 are all the available menu tabs. In the informations menu in figure 22(a) users can open up a page for all the galaxies that have been observed from the current target. The page shows detailed spectra of the galaxy in question with a large variety of well known spectral lines included in the plot. It is possible, but generally not necessary to change the redshift of the galaxies and set notes and other details. Also available is a link to the legacy survey interactive images. In the cluster redshift menu in figure 22(b), the user can assign the cluster redshift to be equal to any member galaxy redshift, to use redMaPPer redshift or the user can set one manually. By default it is the bi-weight average of the member galaxy redshifts. In the cluster center menu, center coordinates can be assigned to the optical center, which is the brightest cluster

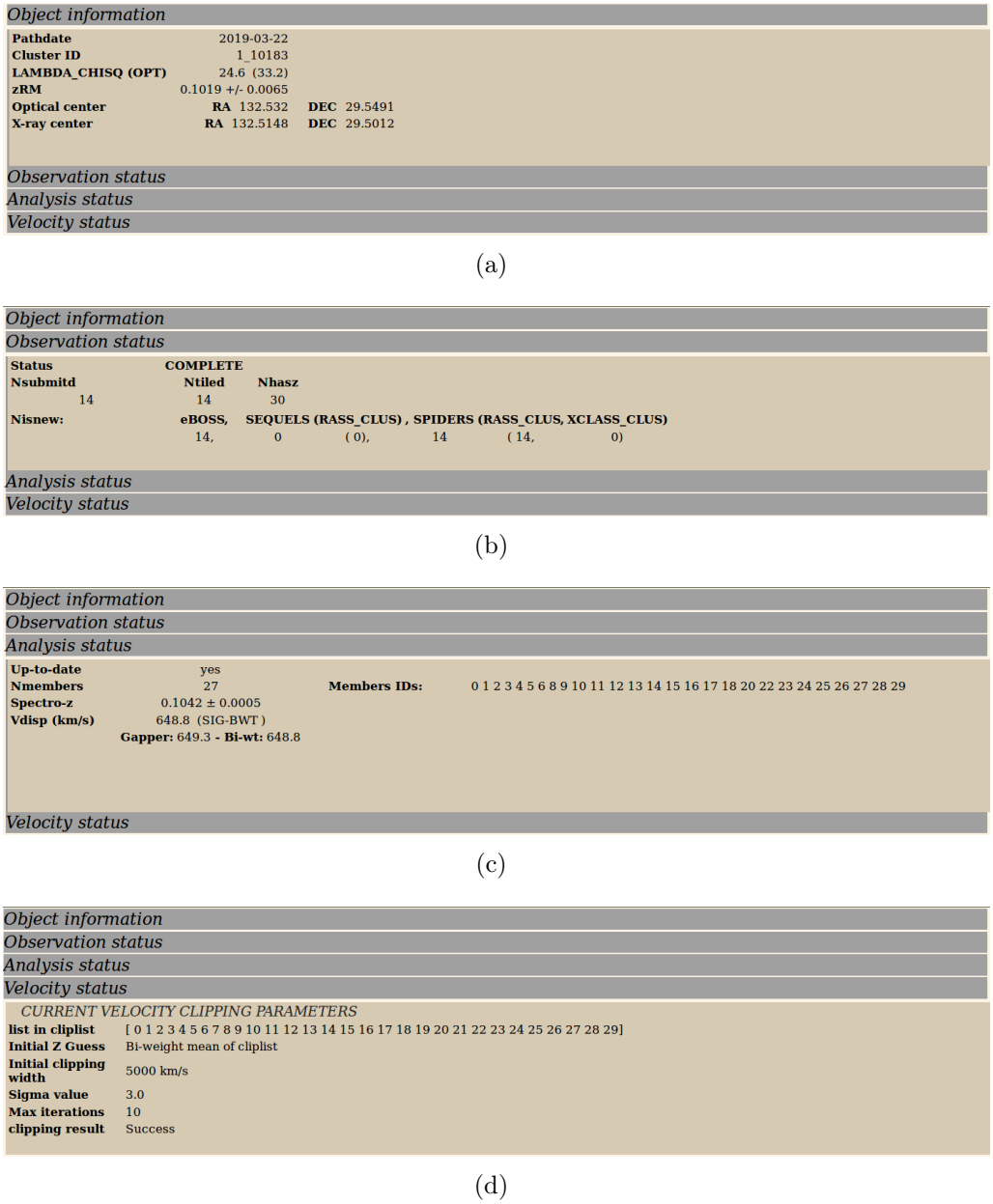


Figure 21: The SPIDERS screening web interface information menus for the example cluster 1_10183: (a) information about the cluster; (b) observation status; (c) analysis status; and, (d) velocity status of the cluster

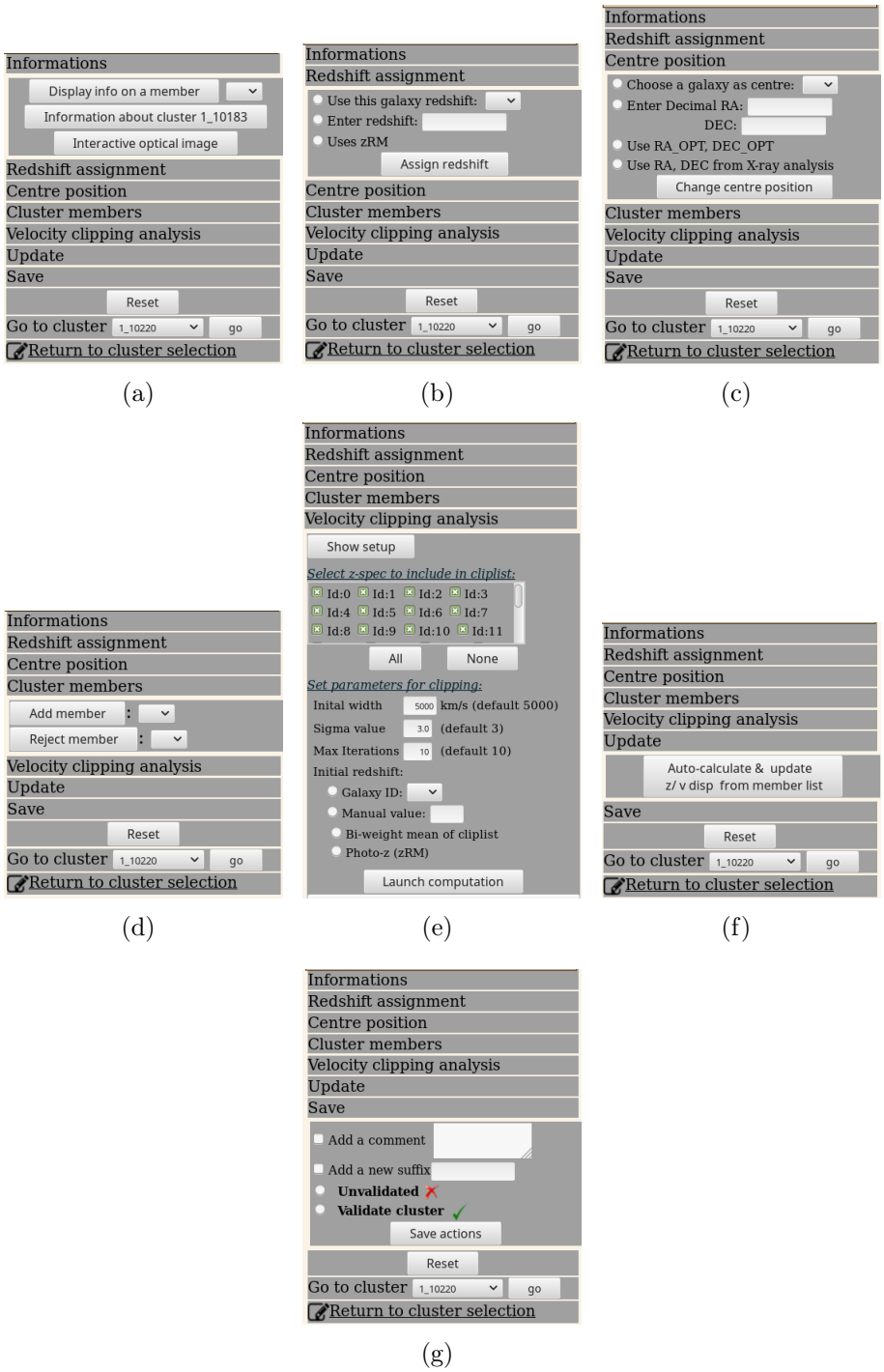
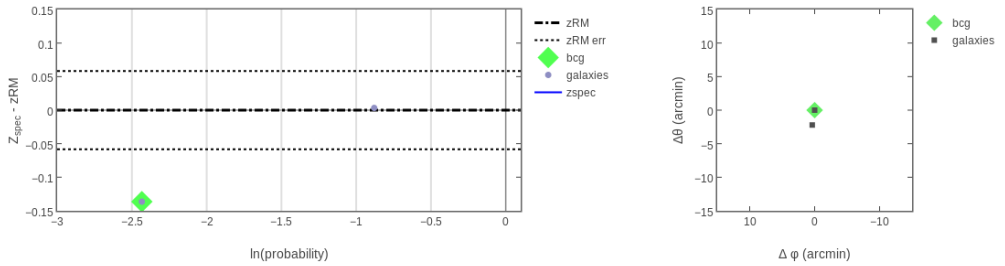


Figure 22: The SPIDERS screening web interface side menus for the example cluster 1_10183: (a) informations about the cluster and selected members; (b) redshift assignment for the cluster; (c) center position for the cluster; (d) cluster members; (e) velocity clipping analysis; (f) update; and, (g) save

galaxy assigned from the pipeline, to the x-ray center, to any member galaxy or the user can set coordinates manually. Default is the optical center. Cluster member galaxies can easily be added or rejected from a list of all available galaxies in the target in the cluster members menu pictured in figure 22(d). The visual screening web interface has an option to select members for velocity clipping analysis, as can be seen in figure 22(e). The user can include members manually or use different parameters which consist of initial width (default value 5000 km/s), σ value (default value is 3.0) and maximum iterations (default value is 10). Initial redshift can be assigned by selecting a galaxy, bi-weight average of clip list, redMaPPer photometric redshift or one can be set manually. After setting desired parameters and clip list, the user can launch computation or restore initial values. After any changes, the cluster properties must be updated from the update menu in figure 22(f). When the user comes to a decision, it is time to save results. The save menu is shown in figure 22(g). If the user has determined that the sample in question is a valid cluster and the members chosen belong to it, the user will choose "validate" from the save tab and save the results. Sometimes the cluster is a valid cluster, but there may be multiple components. The redMaPPer pipeline may have produced a set of galaxies which are part of different close by clusters or there may be contamination from galaxy groups. In this case, the user can choose to split the cluster into two or more components. This can be done by writing a suffix to a text box and saving the results separately. If the components are roughly equal: within 25 percent of each others member count, the chosen suffix is "EQUAL" with a index number after it. If there seems to be a cluster and contamination from groups, then the chosen suffixes are "MAIN" for the main component and "PROJ" for projection effects, again with an index number attached if necessary to the projected objects. We made a general guideline to only split clusters if they all are clear cases. A good example of splitting an object into two components is give in figures 26 and 27 where there is a clear gap between the two components. The gap can be seen in the plots where velocity offsets are visible: 26(a), 26(c) and 26(f). After splitting the object from one to two components the calculated velocity dispersion drops severely. Before splitting the velocity dispersion is $\sigma = 1493.7$ km/s and after splitting, for the main component it is $\sigma = 383.9$ km/s. The redshift of the cluster also changes, but not so dramatically from $z = 0.4578 \pm 0.0038$ to $z = 0.4593 \pm 0.0007$.

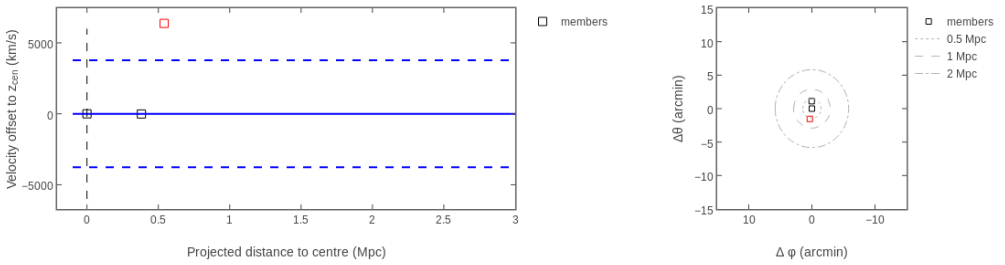
The targets are not all validated, but can also be un-validated by the user. This happens mostly due to insufficient observations ie. there are not enough galaxies within reasonable velocity space and sky positions from each other. The minimum number of members, for a cluster (or component) to be validated, is 3. In figure 23, there are only 2 galaxy spectra in the member selection, so it cannot be validated. In some cases there may be 3 or more galaxies in the target selection, but their velocity offsets are very large, for example 4000 km / s from the cluster center value, in which case the confidence for the target to be a real cluster drops. Such a case is in figure 24.

At rare occasions the automatic pipeline fails to find a cluster, even if it is possible to validate it. One example of a case like this is in figure 25, where there are 3 members



(a) (b)

Figure 23: Example of a un-validated cluster 1_10403. With only two members observed, this object does not qualify to be saved as a validated cluster.

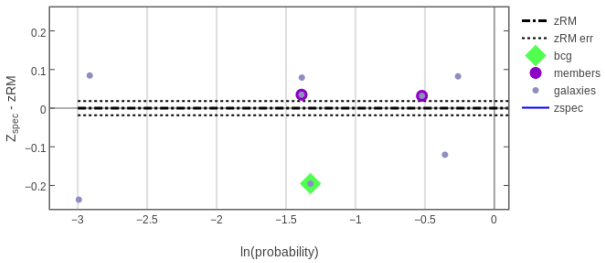


(a) (b)

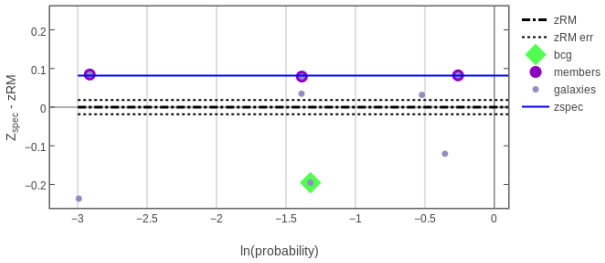
Figure 24: Example of a un-validated cluster with id 1_11922. This object has 3 members observed, but only two are in reasonable velocity distances from each other to be considered in belonging to the same cluster.

in a reasonable velocity-distance space from each other. The pipeline failed in the allocated number of iterations to converge on the actual target. In this case the user can pick out the necessary members and validate the cluster.

Every user validating clusters has their own personal process of decision making. Here I describe the main aspects of the validation process in a general manner. The membership probability plot is useful for showing velocity offsets from the redMaPPer redshift for the object, but also it shows well the individual galaxies' offsets from the cluster and especially it shows well if there are multiple components in the sample, if they are sufficiently far from each other. See example in figure 26. What is sufficiently far, depends on the scale of the plot, but it is in the area of redshift separation of 0.01. The "projected distance to centre (Mpc)" plot shows initially the suggested cluster from the pipeline, where maximum velocity offsets from bi-weight centre are 5000 km/s. Components can also be seen in this plot if they have a clear velocity offset gap between them. There are some cases



(a)



(b)

Figure 25: Example of a automated pipeline failure of cluster with id 1_7452 in 25(a). This object can be validated, because there are 3 members in a reasonable velocity-distance space. The user must pick out the 3 members and validate the object. The redshift of the validated cluster in 25(b) becomes $z = 0.5757 \pm 0.0019$ and the velocity dispersion becomes $\sigma = 596.8$ km/s.

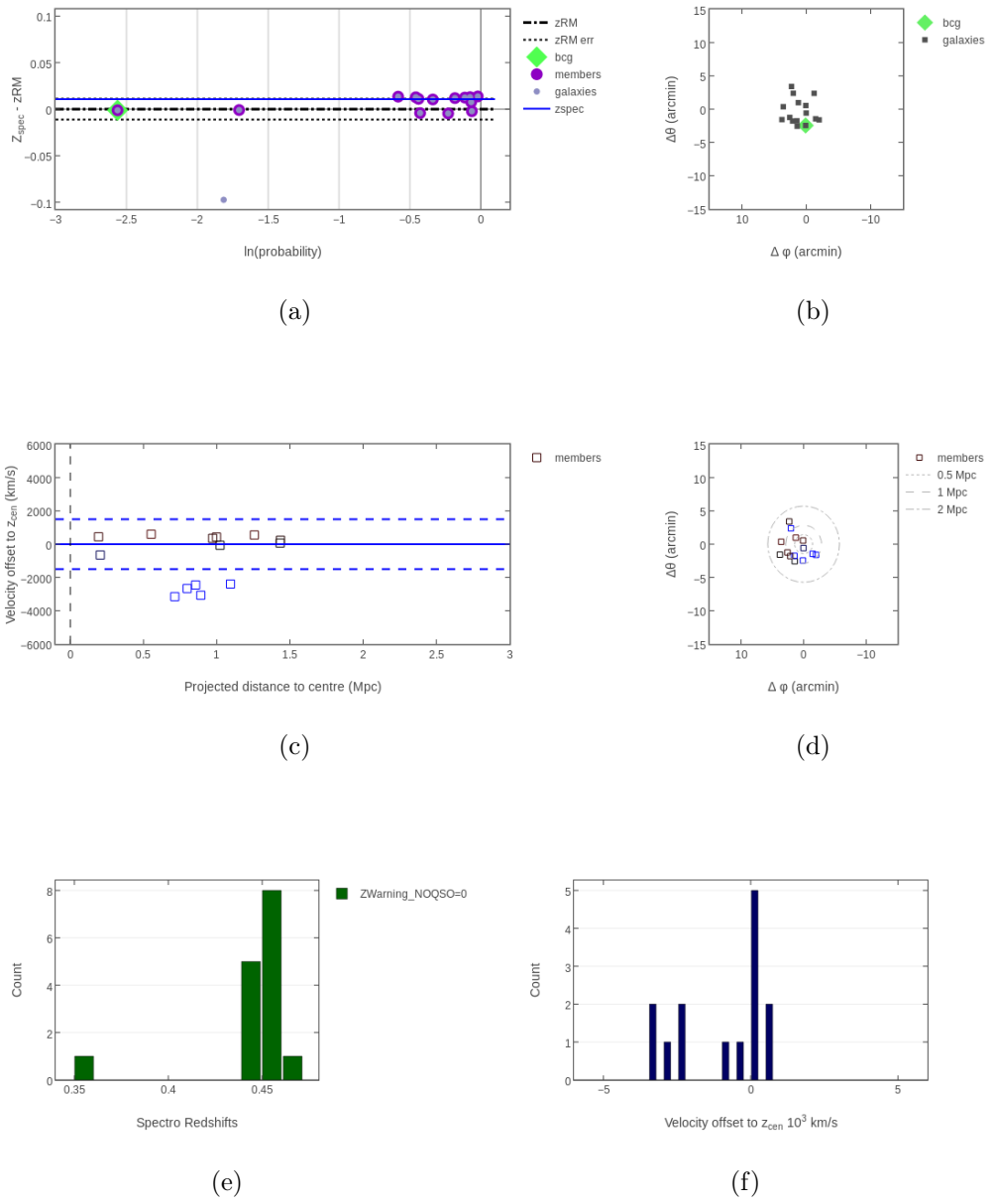
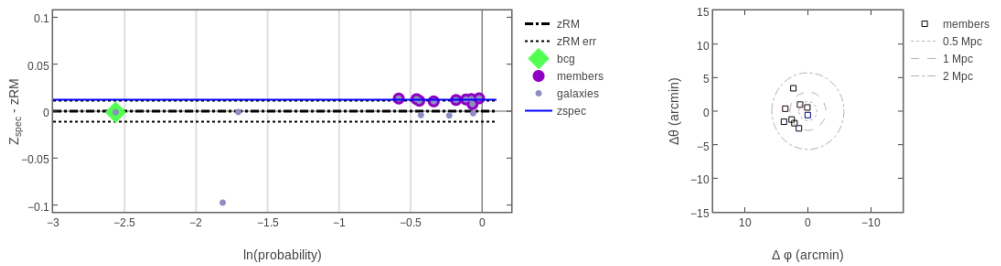
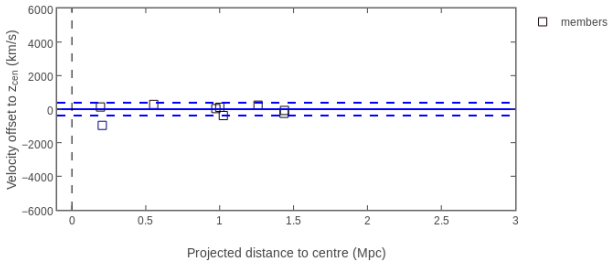


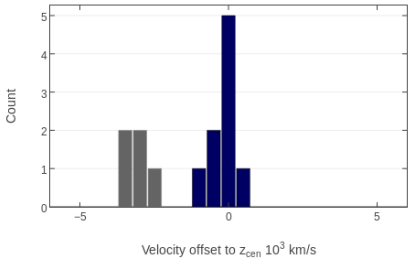
Figure 26: The SPIDERS screening web interface plots for galaxy cluster 1_10157. This object has two components, which the automatic pipeline fails to catch, because they are so close to each other.The solid blue vertical lines in plots 26(a) and 26(c) show the bi-weight average redshift for the combined object, but in both plots a gap between galaxies is visible. Looking at histogram 26(f) the gap is even more clear to see. The redshift for this combined object is $z = 0.4578 \pm 0.0038$ and the velocity dispersion is very high: $\sigma = 1493.7$ km/s.



(a) (b)



(c)



(d)

Figure 27: The SPIDERS screening web interface plots for galaxy cluster 1_10157 after selecting only the main component. The redshift is now $z = 0.4593 \pm 0.0007$ and the velocity dispersion is $\sigma = 383.9$ km/s. Two of the six available plots are not repeated as they are the same as in figure 26

where there may be interactions or even mergers between the components and the users must decide if these should be classified as different components or one. In these cases it is useful to look at the "velocity offset from z_{cen} " histogram. In this plot the user can see the overall state of velocities in the object and whether the object follows a Gaussian curve, which is usually the defining factor of a relaxed cluster. A good example of a velocity histogram following a Gaussian is seen in figure 20(f). Gaussianity is a sign of dynamical equilibrium and it is not a requirement for validation, but it helps the user with decision making. Multiple components should in general have a clear gap in the histogram. If the gap is not clear, the object should not be split into components. There are however many aspects to consider especially with high redshift and/or poorly sampled objects. In the objects with only a few galaxies observed, the users must decide whether the objects can be validated at all (see figure 25). In these cases the user can look at the "velocity offset from centre (Mpc)" plot to see if the galaxies are close to each other in velocity space and in the second sky position plot if they are close to each other in sky coordinates. If neither is true then the object is likely to be un-validated. Richness plays a significant role in the decision making. Rich clusters are expected to have wider caustic mass profiles than smaller richness clusters due to their overall mass differences. If there is a large difference in richness's at the optical centre versus the X-ray centre, it can be an indicator of contamination (all though ROSAT data is not very precise due to its large point spread function). Understanding caustic mass profiles for a range of optical richness is very helpful. Users must use their own knowledge and insight to decide how to treat every object, especially when the member counts are small at higher redshifts where the spectra available becomes scarce due to observational effects.

After all inspectors have finished their validations, it is time to put together the results of the evaluation period. All inspectors have equal votes and a majority must agree on the status of each candidate cluster - either validated or un-validated. The redshift values need to agree within their 95% uncertainty range. The validations are gone through by a moderator, who will run a script to detect disagreements between the votes and also find if inspectors split a cluster into multiple components. If there is no consensus, then those targets are put to the side and wait for more inspectors to vote on them. After all the candidates have gotten a majority vote, there is a conciliation round, where statistics for each cluster are put together. Inspectors may have chosen different members for the clusters, so each galaxy is given a weight according to how many inspectors chose it to belong in the cluster. All redshifts are then averaged as well as velocity dispersions and uncertainties. A redshift spread is computed as the standard deviation of the inspectors' cluster redshift values, where the amplitude depends on the inspectors agreement on membership galaxies. [32]

Figure 28 shows all the spectroscopically validated (black dots) and un-validated (yellow dots) candidate clusters from all the evaluation runs that were done. The figure is plotted with spectroscopic redshift $z_{\lambda,OPT}$ on the x-axis and redMaPPer determined photometric richness λ_{OPT} on the y-axis. The red line following $\lambda_{OPT} = 15 * \zeta(z_\lambda)$ presents roughly the limit where clusters have enough observed member

galaxies to be validated. $\zeta(z_\lambda)$ is a redshift dependent photometric depth correction factor for richness estimation which relates to the redMaPPer SDSS data magnitude cutoff at m_*+ 0.75 (see equation 8 for definition of m_*). The difference between the plotted properties of the validated and un-validated candidate clusters is quite evident from the figure as the majority of low redshift clusters were validated on the left side of the red line even at low richness, and at higher redshift on the right side of the red line only high richness clusters got validated status. [29][32]

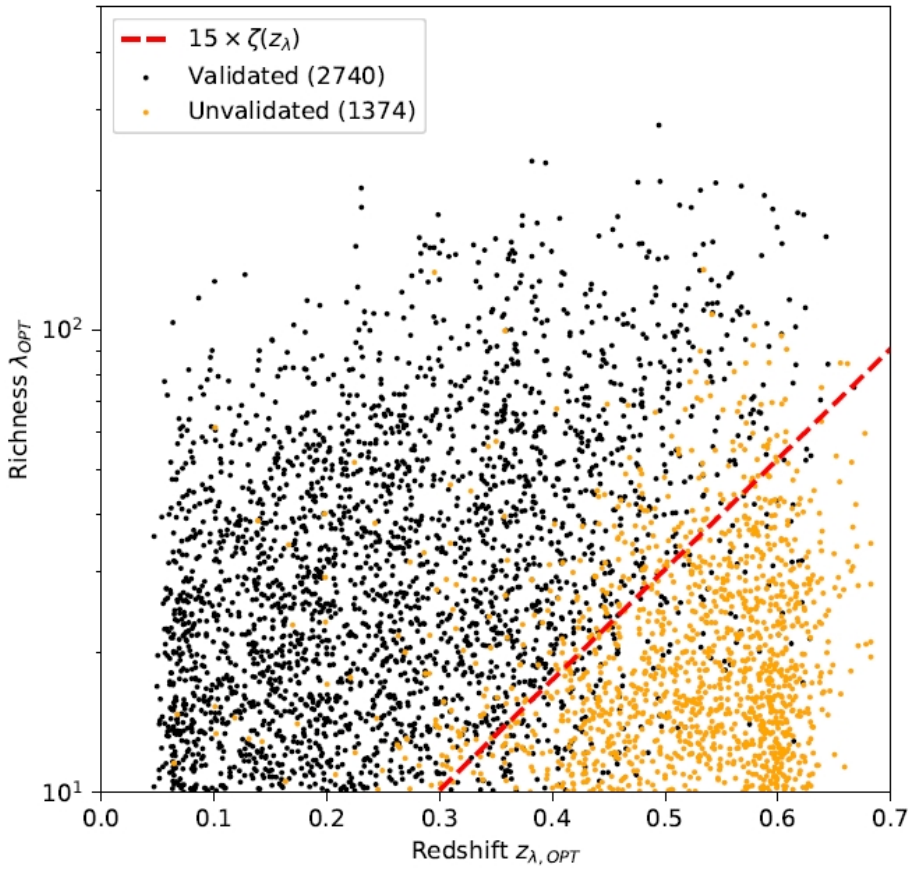


Figure 28: Validated and un-validated candidate clusters richness and redshift space. The black dots are validated clusters and the yellow dots un-validated clusters. The red line is the rough divider for the two possible statuses. See text for the details of the line. [32]

4.3 Inspection runs for the spectroscopic data

Spectroscopic data was gathered over the range of SDSS-IV. There were several data collection periods. These data sets were then evaluated in several different runs during the time of the project. The first two runs were quite large so they were divided into rounds according to their richness at the optical center to make the data more manageable. The cuts for the rounds were $10 < \lambda_{OPT} < 20, 20 < \lambda_{OPT} < 30$ and $\lambda_{OPT} > 30$. Only candidate cluster with richness $\lambda_{OPT} > 10$ are considered.

Run	Inspection round	Selection	Candidates	Validated components	Unique components	Inspectors
2016-07-04	Round-1	$\lambda_{\rm OPT} > 30$	573	520	478	8
2016-07-04	Round-2	$20 < \lambda_{\rm OPT} < 30$	344	246	230	8
2016-07-04	Round-3	$10 < \lambda_{\rm OPT} < 20$	618	396	344	2
2018-04-27	Round-1	$\lambda_{\rm OPT} > 30$	434	379	365	6
2018-04-27	Round-2	$20 < \lambda_{\rm OPT} < 30$	216	166	145	4
2018-04-27	Round-3	$10 < \lambda_{\rm OPT} < 20$	432	281	264	5
2018-04-27	Round-4	$\lambda_{\rm OPT} > 10$, incomplete (SEQUELS)	193	175	159	2
2018-12-04	Round-1	$\lambda_{\rm OPT} > 10$, w/ revised redshift	383	299	276	4
2019-03-22	Round-1	$\lambda_{\rm OPT} > 10$	818	616	571	3
2019-03-22	Round-2	$\lambda_{\rm OPT} > 10$, remaining partially observed	140	101	87	3
2019-03-22	Round-3	re-inspection required	14	14	14	2

Figure 29: All evaluation runs that were done for the SPIDERS catalogue. The run date indicates when the last observations were done for that run. Runs were divided into rounds for easier and smoother handling of the data. Selection column shows the details for each round. Candidate numbers show both the validated and unvalidated cluster candidates. Validated components show all the validated clusters including multiple component targets. Unique components mean the number of validated clusters. The number of inspector available for each round is shown in the last column. Inspectors may not have evaluate the whole rounds, but every candidate was inspected by a minimum of two inspectors. [32]

Furthermore to make the validation process smooth and as even as possible, each run was divided into chunks of 50 candidate clusters, which the pool of inspectors could choose to assign themselves. Each object was required to have a minimum of two inspectors with no maximum number. Each inspector could decide how many of the 50 object chunks they wish to inspect. In figure 29, are the statistics for each inspection run. The dates are the names of the runs, where the date means when the last observations were gathered for that run. For run 2018-04-27 the fourth round had data from a previous work done in 2016 that were flagged as incomplete initially, but were not going to be observed more due to changes in the survey strategy and targeting. Run 2018-12-04 had, besides normally gathered data, also clusters that needed re-inspection because the spectra processing pipeline changed a little. All previously inspected clusters were run with the new pipeline and clusters which had 1 or more galaxies showing a different redshift, redshift error, or warning flag, were inspected again. The final run 2019-03-22 had all the remaining observations including partial observations, because SDSS-IV was coming to it's conclusion and no more spectra would be obtained. As the final steps before merging all the runs together, multiple component targets were handled so that the component with the highest number of members were chosen as the main contributor for the X-ray emission. In cases with equal number of components, the ones with the lowest redshift were chosen. There were 14 cases where the clusters were so close together in redshift space that they were considered possible merging clusters. These 14 were inspected again in round 3 of run 2019-03-22. The cut between single and multiple component was decided as 4000 km/s deviation from the component centers. Anything closer than this, the clusters were considered as one object. The number of candidates, validated components and unique components are shown in figure 29 along with the number of inspectors for each run. Unique components

mean the number of validated clusters. [32]

4.4 Construction of the SPIDERS catalogue

Construction of the SPIDERS value added catalogue (VAC) begins by concatenating all the observations together. Some clusters were validated more than once during the evaluation runs, because they had new observations or the pipeline changed. The newest evaluation of duplicates was kept and the older ones discarded. redMaPPer was run twice to create the CODEX catalogue and on a few rare cases, a cluster would enter the validation twice with a different ID. To remove these, sky coordinates and redshift were used to filter them out. These few clusters were manually checked to make sure they are really the same clusters, before removing the older duplicate.

The X-ray centers from RASS data have quite poor spacial resolution, only about 2'. For this reason, a decision was made to check all clusters that were within 5 times R_{200} and 8000 km/s of each other to see if they are sharing spectroscopic member galaxies. These clusters were visually inspected and we found out that most do indeed share member galaxies and they were therefore listed as single objects instead of separate ones in the catalogue.

One more check-up was done for the un-validated candidates with $z < 0.3$ and with at least 4 members found in running the automated pipeline. They were inspected again to see if they were indeed correctly un-validated. The same was done with candidates with $z > 0.3$ with at least 8 automated pipeline members. Only 20 were found with these criterion and all were deemed correctly un-validated.

Besides the optical data entering the catalogue, there is X-ray data. Measured fluxes at 0.5-2.0 keV are included. Luminosities are calculated for the 0.1-2.7 keV band within the aperture of the clusters 500 times critical density radius (R_{500c}). R_{500c} is calculated from X-ray luminosity - mass scaling relations (L_X-M) from existing studies. Luminosities can be calculated when the distance to the target is known, which can be estimated with the redshifts assuming a cosmological model. For CODEX the ΛCDM model was used with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_M = 0.3$ and $\Omega_\lambda = 0.7$. CODEX clusters are modelled carefully with a selection function that takes into account properties correlating with redshift and chance identifications of X-ray sources depending on background photons and different targeting scales.

[32][29]

4.5 Substructure in galaxy clusters

Galaxy clusters are often modelled as virialized spheres, but in reality they often have in-fall of galaxies and possibly interactions with other galaxy clusters and groups. In the following sections are introduced two tests that have been used in testing for substructure in clusters and groups. The tests were done with a previous data set of SPIDERS and not the final catalogue due to time constraints of this thesis. The

sample had 854 clusters including multiple components.

4.5.1 Anderson-Darling test

The Anderson-Darling test (A-D test) evaluates whether two distributions match each other. For the test, three variables are calculated. First, is A^2:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)(\ln\Phi(x_i) + \ln(1-\Phi(x_{n+1-i}))) \tag{24}$$

where x_i are the ordered radial velocities of the member galaxies, n is the number of member galaxies and $\Phi(x_i)$ is the distribution the data is compared to. Second variable for the A-D test is A^{2*}:

$$A^{2*} = A^2(1 + \frac{0.75}{n} + \frac{2.25}{n^2}) \tag{25}$$

And finally the significance level α:

$$\alpha = a\exp(-A^{2*}/b) \tag{26}$$

where a = 3.6789468 and b = 0.1749916.

For this thesis the distribution to use is a Gaussian cumulative distribution function:

$$\Phi(x_i) = \frac{1}{2}(1 + \mathrm{erf}(\frac{x_i - \mu}{\sqrt{2}\sigma})) \tag{27}$$

where μ is the mean velocity and σ is the velocity dispersion. I set the limit for the sample to be non-Gaussian as $\alpha < 5\%$. Relaxed galaxy clusters are believed to follow a Gaussian velocity distribution. If the distribution is non-Gaussian, one can assume the cluster may be experiencing interactions with other clusters or groups, or has had a recent in-fall of galaxies. [33]

4.5.2 Dressler Shectman test

The Dressler Shectman test (DS test) evaluates whether a cluster has substructure by taking samples of the observed data and comparing the local velocity dispersions to the cluster velocity dispersion. Big deviations from the cluster velocity dispersion indicate substructure. The samples consist of a galaxy and it's nearest neighbors in sky coordinates. In the original paper of Dressler and Shectman from 1988, they used 10 nearest neighbors so that the local sample had 11 member galaxies. In

more recent studies a number scaling with member galaxy counts is used instead: $N_{nn} = \sqrt{N_{members}}$. Large cluster member counts are useful for the test's accuracy. With data consisting mostly of small numbers of observed member galaxies, only a few of these galaxies will be tracing the underlying substructure. The test calculates first a list of local values with:

$$\delta^2 = (N_{nn}/\sigma^2)[(v_{local} - v_{mean})^2 + (\sigma_{local} - \sigma)^2] \tag{28}$$

where v_{local} and v_{mean} are the local and cluster mean velocities and σ_{local} and σ are the local and cluster velocity dispersions. δ values are used to calculate the actual test variable Δ:

$$\Delta = \sum(\delta)/N_{mem} \tag{29}$$

Originally, a critical value was used to estimate the presence of substructure. If the sum of the delta values divided by the number of cluster members was higher than 1, the cluster was flagged as having substructure. A better approach, used in later studies, is to estimate a probability instead. This probability P is calculated by Monte Carlo shuffling. The number of shuffled samples that give a higher delta value than the real cluster are summed together and divided by the number of Monte Carlo iterations. If the cluster delta value is not easily obtained through shuffling, this indicates that the substructure is real. A cluster with P lower that 5% is considered as having substructure. P is calculated as:

$$P = \sum(\Delta_{shuffled} > \Delta_{observed})/n_{shuffle} \tag{30}$$

Visualizing the DS test is done through so called bubble plots, where the scale of the "bubbles" indicate the amount of local deviation from the cluster velocity dispersion $\exp \delta_i$. The bubbles are located at the spatial coordinates of the member galaxies. Big bubbles close together are indicators of the place of substructure. Figure 30 shows an example of a bubble plot. The colors indicate the δ of each galaxy and are used for visualization purposes only. The Δ for this example cluster is 1.306 and $P = 0.748$, which is not small enough to indicate substructure. The results of the DS-test are discussed in section 5. [34][35]

5 Results

5.1 Value added catalogue

The catalogue was released online as a part of the SDSS-IV data release 16 (DR16). The catalogue is the largest of it's kind, consisting of 2740 visually inspected galaxy clusters with 33340 individual galaxy members in them. The catalogue has the optical properties, redshifts, dynamical properties, and X-ray properties of all the

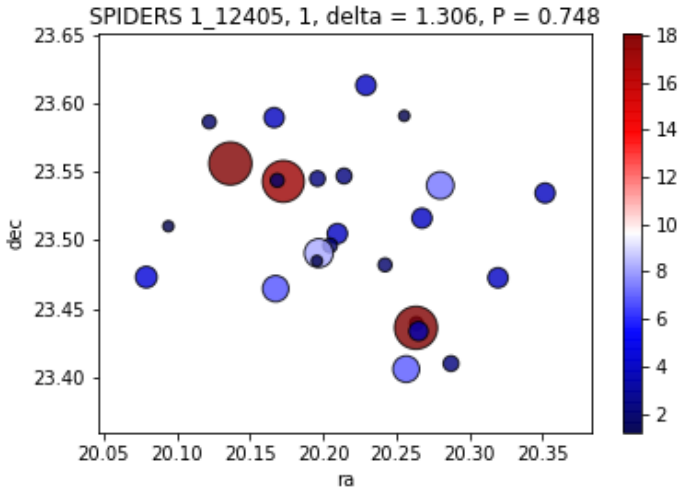


Figure 30: Example of a bubble plot for a SPIDERS cluster with ID 1_12405. The colors indicate the δ of each galaxy and are used for visualization purposes only. The size of the bubbles scales with indication of substructure. The Δ for this example cluster is 1.306 and $P = 0.748$, which strongly indicates no substructure.

validated clusters. The redshift range for the final sample is $0.016 < z < 0.677$. The uncertainty of systems depends on their redshift and typically follows $\Delta_z/(1-z) = 6 * 10^{-4}$. Clusters in the catalogue have member counts between 3 and 75. The median number of galaxies per cluster is 10 and the mean is 12. The mean velocity dispersion for clusters with at least 15 members is 620 km/s. [32][30]

Table 1 has an example of what the catalogue shows for the example cluster 1_10157 from figures 26 and 27. The first column of the table has keywords for the different data. The second column describes the keywords and the third column has the example data. Line 1 shows the SPIDERS ID for the cluster. Line 2 has the number of validated components. Line 3 has the unique CODEX cluster candidate identifier. Lines 4 and 5 have the CODEX X-ray detection right ascension and declination (J2000). Lines 6 and 7 have the CODEX optical cluster center right ascension and declination (J2000). Line 8 has the CODEX Richness λ_{OPT} centered at the optical center. Line 9 has the CODEX photometric redshift z_λ and it's error is on line 10. Line 11 has the number of members in the red sequence and line 12 the number of red sequence members that has gotten spectra **(correct?)**. Line 13 has the cluster spectroscopic redshift after visual inspection and line 14 has the uncertainty of it from bootstrap resampling. Line 15 has the dispersion of redshift from inspections. Lines 16 and 17 have the velocity dispersions from the gapper and bi-weight variance methods. The best velocity dispersion is shown on line 18 (choice of used method depending on the number of validated members). Line 19 shows the number of weighted members after inspection. Lines 20 shows the status of the cluster after inspection. Lines 21 and 22 show the number of inspectors for the candidate and the number of inspectors who validated the candidate as a galaxy cluster. Line 23 has the luminosity in 0.1-2.4 keV band calculated within R_{500c} and

line 24 the uncertainty. Line 25 has the R_{200c} radius of the cluster determined from X-ray luminosity. Lines 26 and 27 have the X-ray flux and it's uncertainty in the 0.5-2.0 keV band. Lines 28 and 29 are for an identifier and alternative name, if such exists in another X-ray survey catalogue called MCXC. [32]

The success rate for identifying CODEX clusters is 100% at redshifts $z < 0.3$ when a minimum of 5 cluster members have spectra from the SPIDERS program. At higher redshifts the depth of the observations becomes the obstacle to fully identify all sources. [29]

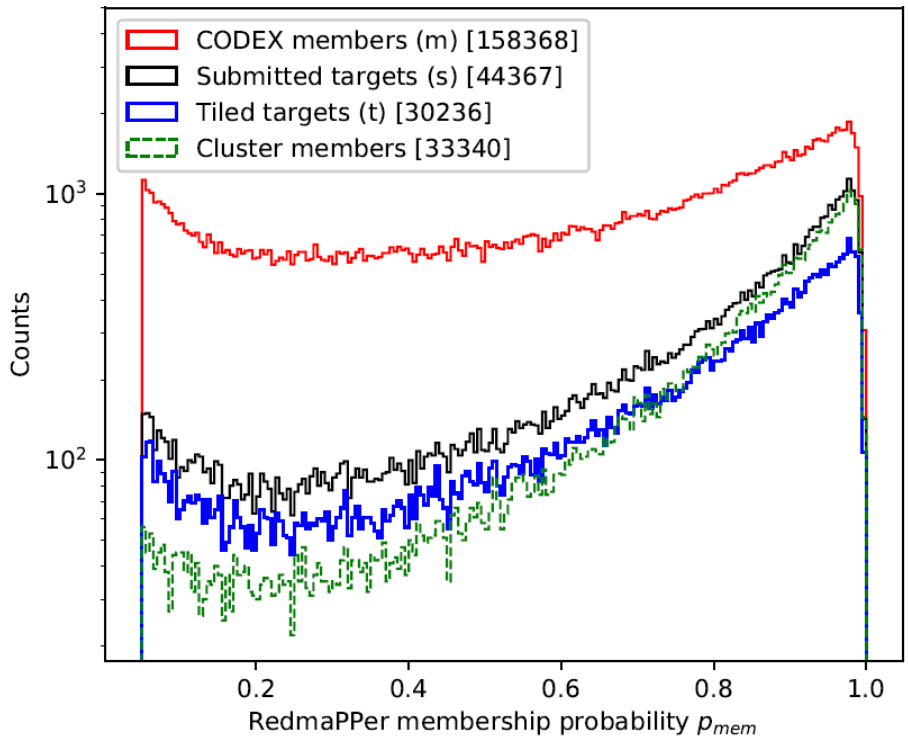


Figure 31: Histogram of targets plotted with redMaPPer membership probabilities. In red is the number of CODEX members, in black all the targets submitted for spectroscopic observations, in blue all the targets which were assigned a fiber and the green dashed line indicates all the final validated spectroscopic members in the SPIDERS catalogue. The final cluster members have spectra also from other sources prior to SPIDERS especially for bright galaxies which tend to reside near the central areas of clusters. This can be seen with the green line rising above the blue line at high membership probabilities. Only targets with $p_{mem} > 0.05$ are considered. Bin width in the histogram is $\Delta p_{mem} = 0.005$. [30]

Figure 31 has a histogram of all CODEX members found from running the redMaPPer algorithm, targets submitted for spectroscopic observations during SDSS-IV, targets that were "tiled", meaning the ones which got assigned a fiber in the fiber plates, and the final validated spectroscopic members. On the x-axis is membership probability from redMaPPer p_{mem} and the number of targets is on the y-axis. Out of the 158368 CODEX members less than a third (44367) were submitted and of

Table 1: Example of the information in the SPIDERS value added catalogue. This example has the same galaxy cluster as in figures 26 and 27.

Column	Description	Example
CLUS_ID	SPIDERS/CODEX identification number	1_10157
NCOMPONENT	Number of validated components	2
CODEX	The CODEX cluster candidate unique identifier	34071
RA	CODEX X-ray detection right ascension (J2000)	128.693
DEC	CODEX X-ray detection declination (J2000)	26.192
RA_OPT	CODEX optical detection right ascension (J2000)	128.671
DEC_OPT	CODEX optical detection declination (J2000)	26.186
LAMBDA_CHISQ_OPT	Richness λ_{OPT} of the CODEX optical detection	106.209
Z_LAMBDA	Photometric redshift (z_λ) of the CODEX optical detection	0.447
Z_LAMBDA_ERR	Uncertainty on z_λ	0.011
NMEM	Number of objects in the CODEX red sequence	91
NOKZ	Number of red-sequence members with a spectroscopic redshift	15
SCREEN_CLUZSPEC	Galaxy cluster redshift, assigned after visual inspection	0.459
SCREEN_CLUZSPEC_ERR	Bootstrap uncertainty on SCREEN_CLUZSPEC	0.00066
SCREEN_CLUZSPEC_SPREAD	Dispersion in the inspection cluster redshifts	0
SCREEN_CLUZSPEC_GAP	Gapper estimate of the cluster velocity dispersion	383.917
SCREEN_CLUVDISP_BWT	Square root of the bi-weight variance velocity dispersion	352.588
SCREEN_CLUVDISP_BEST	Value of the "best velocity dispersion"	383.917
SCREEN_NMEMBERS_W	Weighted number of red-sequence members identified as members	9
STATUS	Validation status of the cluster assigned by the visual inspector	validated
NINSPECTORS	Number of individual inspections for this system	2
NVALID	Number of inspectors validating this system as a galaxy cluster	2
LX0124	Luminosity in the (0.1-2.4) keV band of the cluster, aperture R_{500c}	$3.574{*}10^{44}$
ELX	Uncertainty on LX0124	$1.305{*}10^{44}$
R200C_DEG	Apparent R_{200c} radius of the galaxy cluster	0.0697
FLUX052	Galaxy cluster X-ray flux in the 0.5-2.0 keV band	$3.303{*}10^{-13}$
EFLUX052	Uncertainty on FLUX052	$1.206{*}10^{-13}$
MCXC	Identifier in the MCXC catalogue, if present	n/a
ANAME	Alternative name in, if present	n/a

those roughly 70% were tiled up to membership probability $p_{mem} < 0.8$. Higher probability members are concentrated at the centers of cluster candidates and could

not be assigned fibers as often even with overlapping plates as the fiber collision radius is 62". The total number of tiled targets is 30236. Of those some were still not observed due to technical issues, or they were discarded because they were observed in SDSS-III and were deemed sufficient. The number of validated cluster members in the SPIDERS catalogue is 33340. This number includes spectra from other sources prior to the SPIDERS program, mainly the brightest galaxies that tend to reside at the central regions of galaxy clusters. This effect can be clearly seen in the rise of the green validated members curve above the blue tiled targets curve in the histogram at high membership probabilities. [30]

The quality of the spectra was good throughout the survey with only around 1% of spectra with a low signal-to-noise ratio warning from the automated object identification pipeline. The median signal-to-noise is above 3. The main contribution to spectral quality is the flux of the targets, which means the lowest magnitude targets are where the quality is poorest. The median uncertainty in clusters with a reliable redshift fit from templates is about 20 km/s, which is sufficient since clusters have velocity dispersion of about 100 to 1500 km/s. Less than 1% have uncertainties over 60 km/s and only 5 clusters with uncertainty of more than 100 km/s. From the automated photometric classification pipeline, there were 29277 targets with a reliable non-quasi-stellar template fit. Out of these only 117 were classified as a star at redshift zero after getting spectra and the rest were classified as galaxies. This shows how well the photometric classification works to detect galaxies. [30]

Figure 32 shows the confirmed SPIDERS clusters in comoving coordinates. Galaxy clusters with 5 to 15 spectroscopic members are marked with gray circles and clusters with more than 15 members are marked with yellow circles. The circle size is an indication of cluster X-ray luminosities. The observer in the figure is located at the center and northern and southern skies in the left and right. The dashed circles represent redshifts of z = 0.1, 0.3 and 0.6. It is clear to see that high redshift clusters have much larger circle sizes, which is an observational bias called Malmquist bias where richness correlates with redshift. [30]

One of the scientific results using the SPIDERS catalogue is a power-law fit for the scaling relation between X-ray luminosity L_X and bi-weight velocity dispersion σ_{BWT}. For the scaling relation, only clusters with at least 15 members were used (787 in total). The red curve in figure 33 is done by finding the best values for the constants in equation 31. Uncertainties are calculated with resampling of member galaxies from a denser and very clean galaxy cluster survey called HIFLUGCS. The red dashed lines indicate 1σ uncertainties. The gray diamonds are the raw velocity dispersion estimates and the black crosses are bias estimated values, which have uncertainties from the same model as the 1σ except they depend on the member counts of each cluster. The black line is a fit from the HIFLUGS sample, which is quite close to the scaling relation of the SPIDERS sample. [32][10]

$$\log_{10}\left(\frac{\sigma_{BWT}}{700\text{km/s}}\right) = A + B\,\log_{10}\left(\frac{L_X E(z)^{-1}}{10^{44}\text{erg/s}}\right) \tag{31}$$

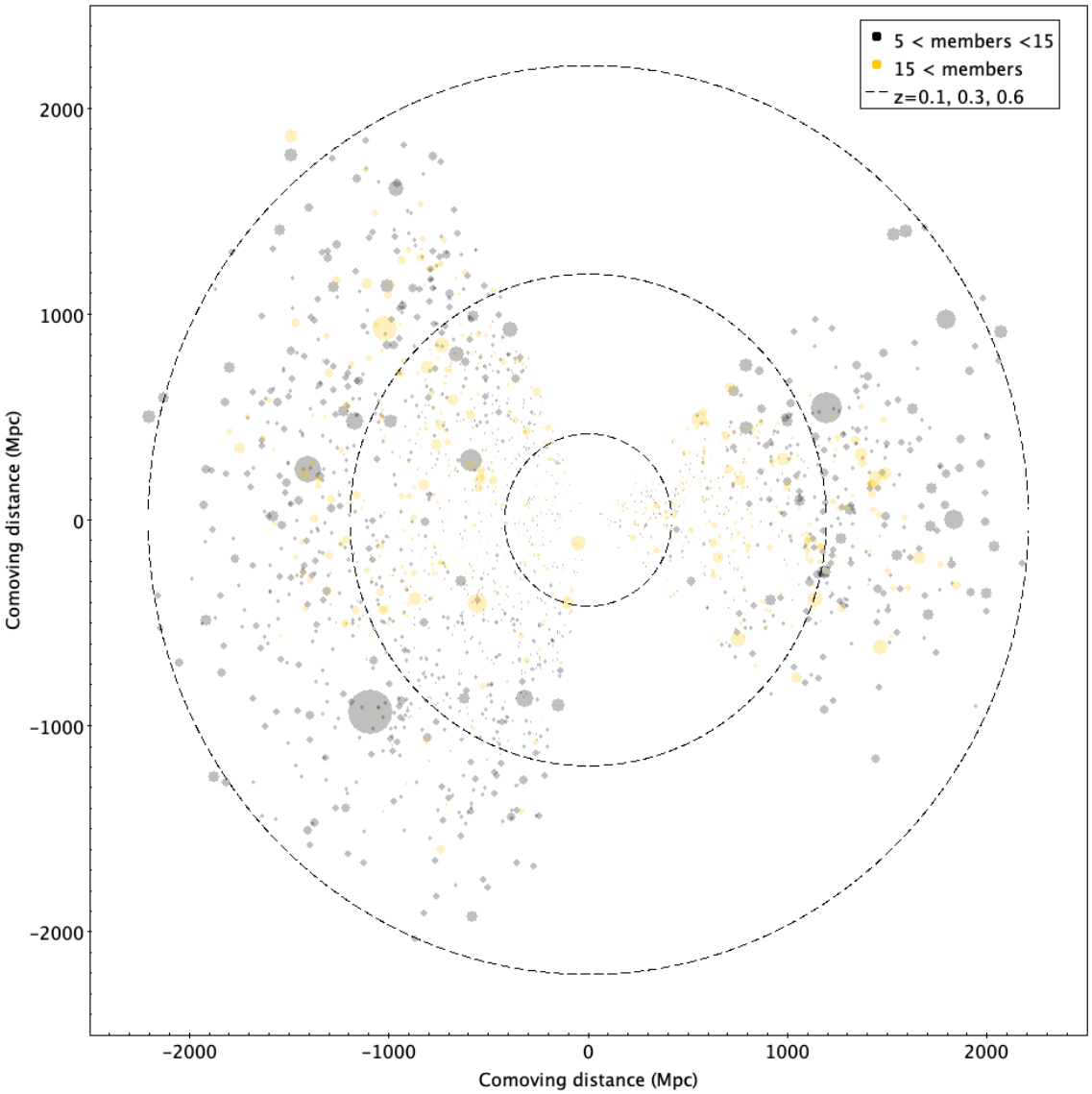


Figure 32: Confirmed SPIDERS clusters at comoving distances. Each colored circle is a cluster. Gray circles are clusters with 5 to 15 spectroscopic members and yellow circles are clusters with more than 15 members. The size of the circles corresponds to X-ray luminosities. The observer is located at the center (0,0) and the northern and southern skies are on the left and right side respectively. The dashed circles indicate redshifts z = 0.1, 0.3 and 0.6. [30]

Figure 34 is the X-ray luminosity function for CODEX/SPIDERS clusters. The dots represent CODEX data and crosses SPIDERS data. The light gray data points are for clusters in the redshift range $0.1 < z < 0.3$ and the black points are for $0.3 < z < 0.6$. The curve shows the Schecter function for another X-ray cluster sample called REFLEX which is a very close to the SPIDERS sample data points. The use of spectroscopic redshifts versus photometric redshifts does not effect the function very much, because of large redshift bins and the slow evolution of X-ray luminosity

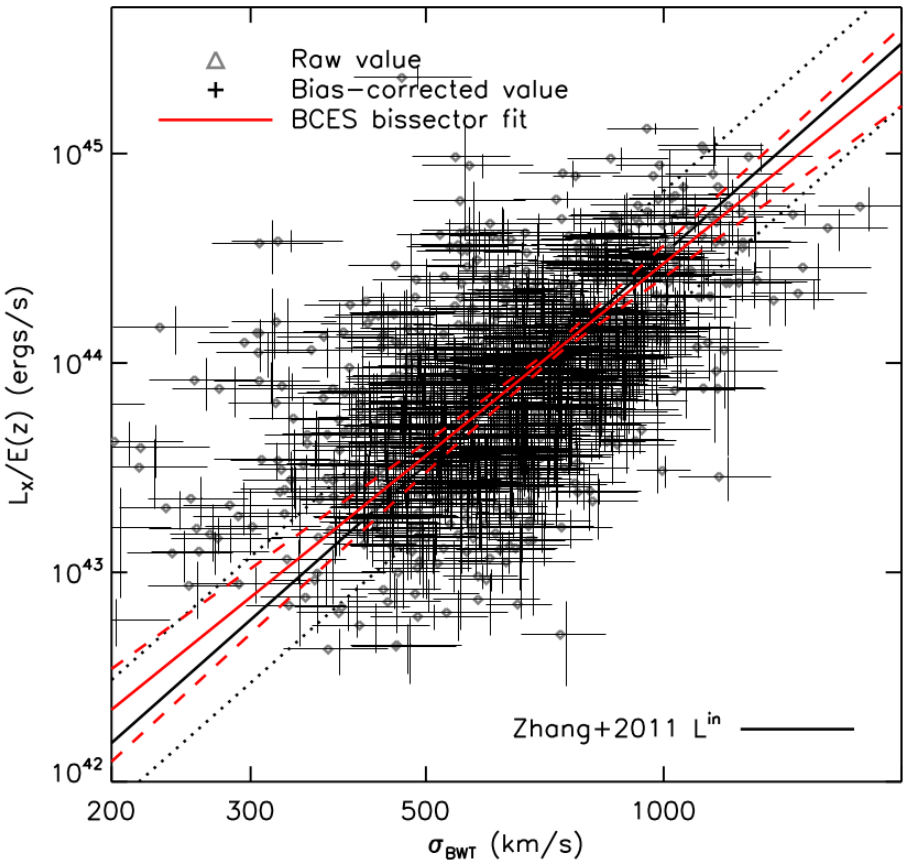


Figure 33: X-ray luminosity - velocity dispersion relation for SPIDERS clusters with at least 15 confirmed member galaxies. The red solid line is the best fit power-law. The red dashed lines indicate 1σ uncertainties. The gray diamonds are the raw velocity dispersion estimates and the black crosses are bias estimated values, where uncertainties have a dependency on cluster member counts. The black line is a fit from another galaxy cluster survey called HIFLUGS. [32]

in clusters. [30]

5.2 Results from the substructure tests

The SPIDERS galaxy cluster catalogue is the main subject of this thesis, but I also did a small side project with an earlier version of SPIDERS data in which I studied how well substructure can be found in the SPIDERS clusters with two different, commonly used substructure tests.

5.2.1 Results of the Anderson-Darling test

I performed the A-D test on a small sample of SPIDERS data consisting of 853 clusters and cluster components. The test has not been done on the final SPIDERS sample due to time constraints on the thesis. Some of the tested clusters had been

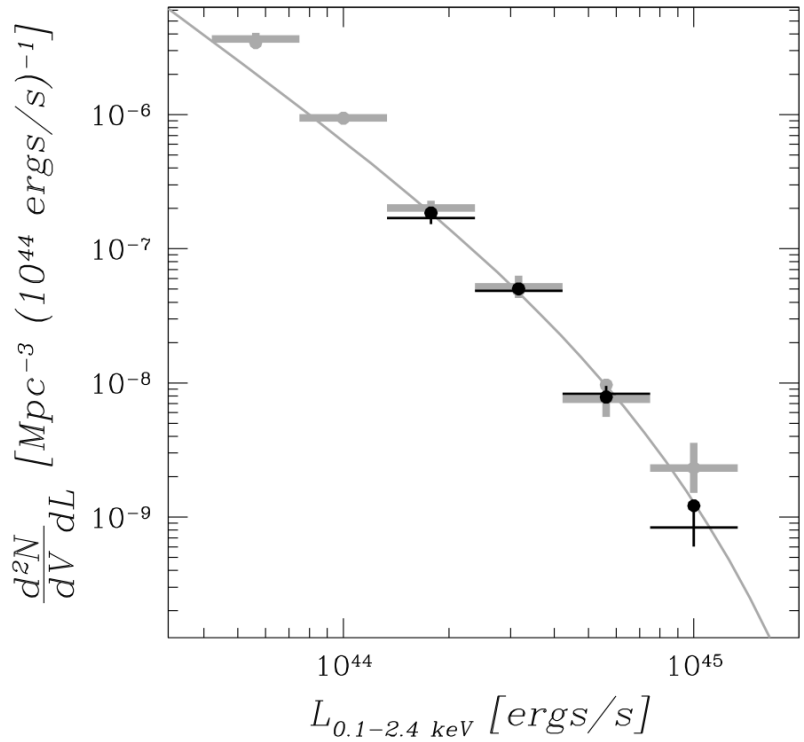


Figure 34: X-ray luminosity function for the SPIDERS clusters. The dots represent CODEX data and crosses SPIDERS data. The light gray data points are for clusters in the redshift range $0.1 < z < 0.3$ and the black points are for $0.3 < z < 0.6$. The curve shows the Schecter function for another X-ray cluster sample called REFLEX [30]

split into multiple cluster components having the same SPIDERS ID. After running the test algorithms, the results had 771 Gaussian flags and 82 non-Gaussian flags. The A-D test has been tested by other scientist in the past (for example Hou et al. 2009), who found it to be a good test for finding substructures in galaxy clusters. The test performs reliably even down to $n = 5$, which is the minimum number of member galaxies I used for the SPIDERS sample as well.

I also looked at how the Gaussian and non-Gaussian clusters differ compared against different properties of galaxy clusters. I found no clear differences when compared with redshifts, X-ray luminosities or richnesses. There is some difference in how the Gaussian and non-Gaussian clusters compare to velocity dispersions of the clusters, as can be seen in the shape of the histograms in figure 35. The Gaussian cluster in this sample peak at velocity dispersion around 600 km/s, when the non-Gaussian clusters have a much shallower peak at a slightly larger velocity dispersion value at 700 km/s. Note that the histogram has a logarithmic scale on the x-axis. This result is not surprising since the in-fall of galaxies and mergers or interactions with clusters and groups of galaxies stir the clusters dynamical state.

These results are an indicator that galaxy clusters are quite often experiencing on-going in-fall of galaxies and interactions from other groups or clusters and the A-D

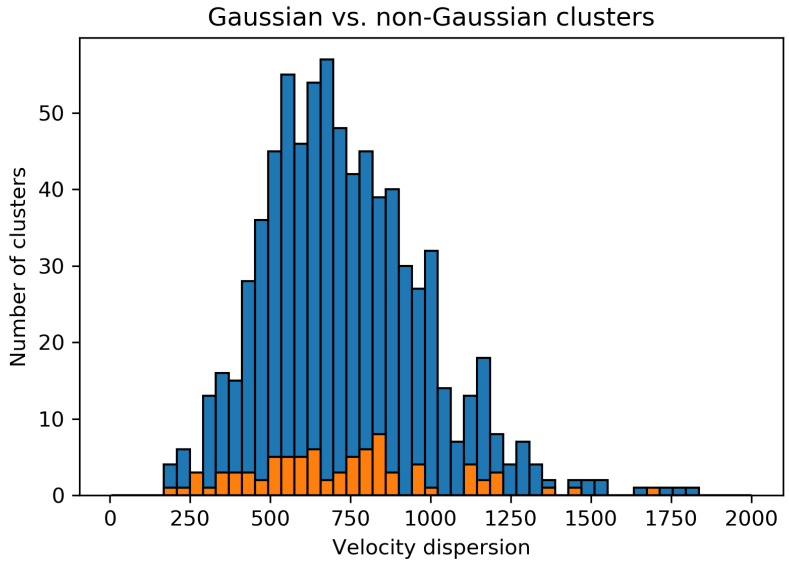


Figure 35: Gaussian and non-Gaussian clusters compared to the velocity dispersion of the galaxy clusters. The A-D test was not done with the final SPIDERS catalogue, but a smaller sub-sample. The test resulted in 771 Gaussian cluster flags marked with blue and 82 non-Gaussian flags marked with orange. The peaks of the different sub-samples have different amplitudes and the non-Gaussian clusters have slightly larger velocity dispersions.

test can pick up on it even with mostly red sequence members. This result needs more investigation however and would also benefit from the A-D test being run on the full final SPIDERS sample.

5.2.2 Results from the Dressler Shectman test

For the DS test I chose to include clusters with at least 20 members. Using the original approach from Dressler and Shectman with Bi-weight velocity dispersions produced results that almost all clusters have substructure. I chose to use the Gapper velocity dispersion method, which produces values of 2 and above. Getting unrealistic amounts of substructure flags for clusters has been an issue in previous studies using critical values. Probability values give a much more reliable result as discussed in section 4.5.2.

I performed the test using 1000 Monte Carlo shuffles on the same small sample of SPIDERS cluster as with the A-D test, but only 171 of those had at least 20 galaxies and were included. I decided to test a few times to see if the results changed and they were slightly different every time with varying number of substructure flags - between 5 and 9 - and also some of the clusters getting a flag changed between the runs. This clearly indicates that the test is not all that reliable for this data. Further investigation would be needed if this test was done again with the SPIDERS

clusters. Also the very small number of substructure flags gotten, even with the small number of tested clusters, suggest that the test is not very useful for a sample like SPIDERS. It seems likely that the way the SPIDERS catalogue is done, is effecting the inclusion of substructure by excluding the galaxies with deviating velocities already at the validation process, and also because the outer regions of clusters, where substructure is more likely to occur, are not included in the targeting scheme of SPIDERS.

6 Discussion and conclusions

The SPIDERS program has been a major project for many scientists and other employees of the SDSS collaboration as part of SDSS-IV and eBOSS. The observations, data handling pipelines, validations and scientific work has continued for several years. The result of the work is the largest X-ray selected, spectroscopic and visually validated catalogue of galaxy clusters to date. The spectroscopic identification of sources was done automatically. Also to help alleviate a major workload from visual inspectors, an automated pipeline for cluster candidates was created, which selected most likely galaxies as cluster members and created helpful visual aids for the inspectors to base their decisions on. Most cluster candidates could therefore be either validated or un-validated with a few clicks of a mouse and only the more difficult cases needed adjustments from inspectors. As the visual screening and cluster validation is done by humans, it may of course be vulnerable to the views and opinions of the inspectors. But we believe that having at least two inspectors for every cluster candidate and the automated cluster pre-inspection, the validation process will be robust enough for such a large survey.

The SPIDERS value added catalogue is a massive accomplishment. Putting in the effort of creating it can be justified with the gained precision from spectroscopic observations and the valuable visual validation, which makes it creditable. Figure 36, shows the difference in redshift between the photometric and the spectroscopic samples. The colors are selected according to the final spectroscopic validated member counts. In blue are the clusters with $3 \leq z < 10$ members, in green for $10 \leq z < 15$ and in red for $z \geq 15$. The number of validated cluster members depends strongly on redshift of the clusters. This is logical, because it is easier to detect smaller galaxies nearby and also the fiber collision distance becomes more excluding to dense areas at higher redshift. As can be seen from the figure, at higher redshift and lower number of detected galaxies, scatter increases, but also at very low redshifts the deviation of photometric redshifts is noticeable. The dip at low redshift is caused by known issues of the redMaPPer algorithm at $z < 0.05$. The figure indicates clearly that spectroscopic observations are indeed important for the accuracy of the redshift estimations especially if the number of detected cluster members are low or the cluster is at redshifts higher than approximately 0.2. The small histograms inside the figure show the number of clusters per redshift bins of $\Delta z = 0.04$. The typical improvement on uncertainties, going from photometric redshift to spectroscopic, is

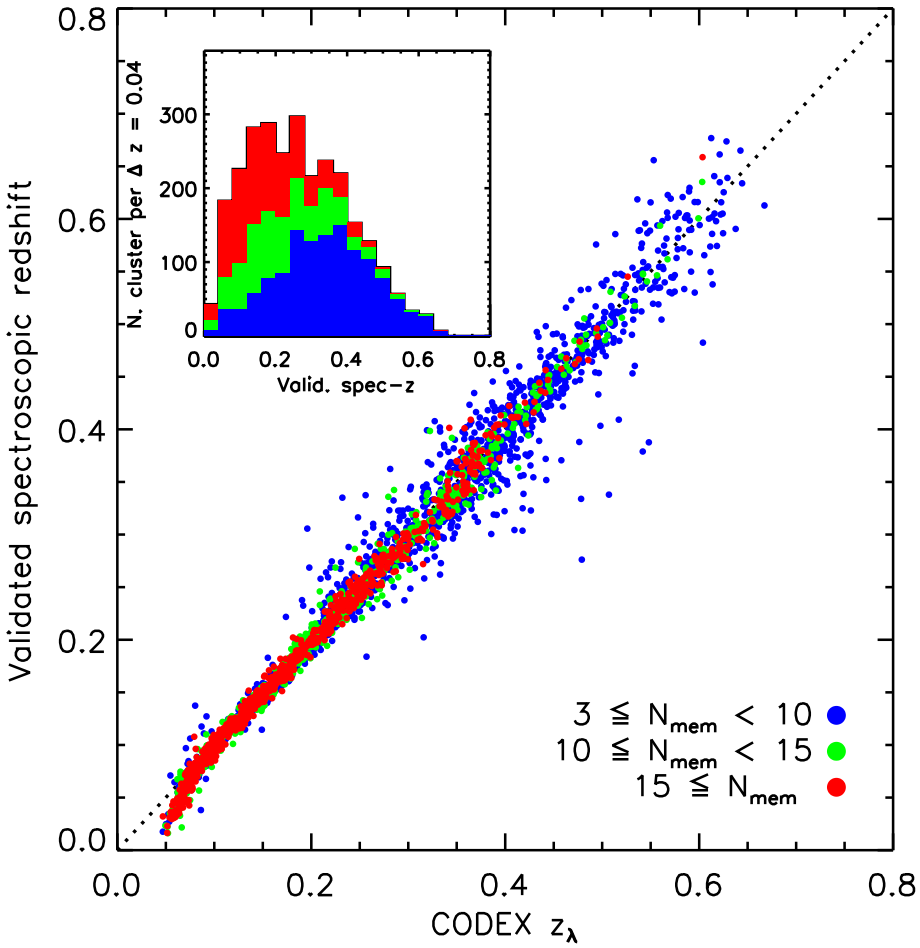


Figure 36: The difference in redshift between the photometric and the spectroscopic samples. X-axis has the photometric redshifts from redMaPPer and the y-axis has the SPIDERS validated spectroscopic redshifts. The colors indicate the final spectroscopic validated member counts. In blue are the clusters with $3 \leq z < 10$ members, in green for $10 \leq z < 15$ and in red for $z \geq 15$. The small histograms inside the figure show the number of clusters per redshift bins of $\Delta z = 0.04$. [32]

about a factor of 10.

For the purpose of constraining cosmological properties and studying large scale structures of the Universe with galaxy clusters, the typically needed qualities for catalogues are large numbers, precise observables, a wide range of redshifts and halo masses and well understood selection effects. All these qualities are achieved with the SPIDERS value added catalogue. The catalogue has already been used for science, for example to define richness-mass and luminosity-mass relations for galaxy clusters and for further defining properties of the brightest central galaxies (BCGs) residing in the clusters. Now that the final catalogue it has been published, it is available for all scientists to use on their own research. Research also continues

within the SPIDERS group of astrophysicists. In other future projects it is also possible to use the results of the SPIDERS program to further develop algorithms and machine learning protocols for galaxy cluster surveys.

The cluster validation process was my personal main contribution to the SPIDERS catalogue. I was one of the inspectors for almost all candidate clusters. This project has taught me a lot about how to conduct an observational survey, how the validation process of such a survey works and most importantly, some valuable aspects of what is needed to be a researcher in observational astrophysics. I also became more confident in programming with python as part of the research and of course learned a lot more about galaxy clusters and especially the on observational side. I made connections to many knowledgeable colleagues in the same field and will continue my personal research later on with the help of this project.

In this thesis I have explained with detail, the process of the SPIDERS program: large scale spectroscopic survey of galaxy clusters. I have shown what is needed to know beforehand, what kind of steps are done in each phase of the survey and explained the methods and algorithms used in the process. The end result is the largest spectroscopic, visually validated catalogue of X-ray selected galaxy clusters to date.

The SPIDERS catalogue can and already has been used for precise and meaningful scientific research in the field of galaxy cluster astronomy. It is a powerful tool for constraining cosmological parameters and studying various properties and evolution of large scale structures of the Universe.

The value added catalogue can be found online at: `https://www.sdss.org/dr16/data_access/value-added-catalogues/?vac_id=spiders-x-ray-galaxy-cluster-catalogue-for-dr16`

References

1 A. V. Kravtsov and S. Borgani, "Formation of galaxy clusters," *Annual Review of Astronomy and Astrophysics*, vol. 50, no. 1, pp. 353–409, 2012.

2 G. M. Voit, "Tracing cosmic evolution with clusters of galaxies," *Rev. Mod. Phys.*, vol. 77, pp. 207–258, Apr 2005.

3 Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., Bartlett, J. G., Bartolo, N., Battaner, E., Battye, R., Benabed, K., Benoît, A., Benoit-Lévy, A., Bernard, J.-P., Bersanelli, M., Bielewicz, P., Bock, J. J., Bonaldi, A., Bonavera, L., Bond, J. R., Borrill, J., Bouchet, F. R., Boulanger, F., Bucher, M., Burigana, C., Butler, R. C., Calabrese, E., Cardoso, J.-F., Catalano, A., Challinor, A., Chamballu, A., Chary, R.-R., Chiang, H. C., Chluba, J., Christensen, P. R., Church, S., Clements, D. L., Colombi, S., Colombo, L. P. L., Combet, C., Coulais, A., Crill, B. P., Curto, A., Cuttaia, F., Danese, L., Davies, R. D., Davis,

R. J., de Bernardis, P., de Rosa, A., de Zotti, G., Delabrouille, J., Désert, F.-X., Di Valentino, E., Dickinson, C., Diego, J. M., Dolag, K., Dole, H., Donzelli, S., Doré, O., Douspis, M., Ducout, A., Dunkley, J., Dupac, X., Efstathiou, G., Elsner, F., Enßlin, T. A., Eriksen, H. K., Farhang, M., Fergusson, J., Finelli, F., Forni, O., Frailis, M., Fraisse, A. A., Franceschi, E., Frejsel, A., Galeotta, S., Galli, S., Ganga, K., Gauthier, C., Gerbino, M., Ghosh, T., Giard, M., Giraud-Héraud, Y., Giusarma, E., Gjerløw, E., González-Nuevo, J., Górski, K. M., Gratton, S., Gregorio, A., Gruppuso, A., Gudmundsson, J. E., Hamann, J., Hansen, F. K., Hanson, D., Harrison, D. L., Helou, G., Henrot-Versillé, S., Hernández-Monteagudo, C., Herranz, D., Hildebrandt, S. R., Hivon, E., Hobson, M., Holmes, W. A., Hornstrup, A., Hovest, W., Huang, Z., Huffenberger, K. M., Hurier, G., Jaffe, A. H., Jaffe, T. R., Jones, W. C., Juvela, M., Keihänen, E., Keskitalo, R., Kisner, T. S., Kneissl, R., Knoche, J., Knox, L., Kunz, M., Kurki-Suonio, H., Lagache, G., Lähteenmäki, A., Lamarre, J.-M., Lasenby, A., Lattanzi, M., Lawrence, C. R., Leahy, J. P., Leonardi, R., Lesgourgues, J., Levrier, F., Lewis, A., Liguori, M., Lilje, P. B., Linden-Vørnle, M., López-Caniego, M., Lubin, P. M., Macías-Pérez, J. F., Maggio, G., Maino, D., Mandolesi, N., Mangilli, A., Marchini, A., Maris, M., Martin, P. G., Martinelli, M., Martínez-González, E., Masi, S., Matarrese, S., McGehee, P., Meinhold, P. R., Melchiorri, A., Melin, J.-B., Mendes, L., Mennella, A., Migliaccio, M., Millea, M., Mitra, S., Miville-Deschênes, M.-A., Moneti, A., Montier, L., Morgante, G., Mortlock, D., Moss, A., Munshi, D., Murphy, J. A., Naselsky, P., Nati, F., Natoli, P., Netterfield, C. B., Nørgaard-Nielsen, H. U., Noviello, F., Novikov, D., Novikov, I., Oxborrow, C. A., Paci, F., Pagano, L., Pajot, F., Paladini, R., Paoletti, D., Partridge, B., Pasian, F., Patanchon, G., Pearson, T. J., Perdereau, O., Perotto, L., Perrotta, F., Pettorino, V., Piacentini, F., Piat, M., Pierpaoli, E., Pietrobon, D., Plaszczynski, S., Pointecouteau, E., Polenta, G., Popa, L., Pratt, G. W., Prézeau, G., Prunet, S., Puget, J.-L., Rachen, J. P., Reach, W. T., Rebolo, R., Reinecke, M., Remazeilles, M., Renault, C., Renzi, A., Ristorcelli, I., Rocha, G., Rosset, C., Rossetti, M., Roudier, G., Rouillé d´Orfeuil, B., Rowan-Robinson, M., Rubiño-Martín, J. A., Rusholme, B., Said, N., Salvatelli, V., Salvati, L., Sandri, M., Santos, D., Savelainen, M., Savini, G., Scott, D., Seiffert, M. D., Serra, P., Shellard, E. P. S., Spencer, L. D., Spinelli, M., Stolyarov, V., Stompor, R., Sudiwala, R., Sunyaev, R., Sutton, D., Suur-Uski, A.-S., Sygnet, J.-F., Tauber, J. A., Terenzi, L., Toffolatti, L., Tomasi, M., Tristram, M., Trombetti, T., Tucci, M., Tuovinen, J., Türler, M., Umana, G., Valenziano, L., Valiviita, J., Van Tent, F., Vielva, P., Villa, F., Wade, L. A., Wandelt, B. D., Wehus, I. K., White, M., White, S. D. M., Wilkinson, A., Yvon, D., Zacchei, A., and Zonca, A., "Planck 2015 results - xiii. cosmological parameters," *A&A*, vol. 594, p. A13, 2016.

4 R. Capasso, A. Saro, J. J. Mohr, A. Biviano, S. Bocquet, V. Strazzullo, S. Grandis, D. E. Applegate, M. B. Bayliss, B. A. Benson, L. E. Bleem, M. Brodwin, E. Bulbul, J. E. Carlstrom, I. Chiu, J. P. Dietrich, N. Gupta, T. de Haan, J. Hlavacek-Larrondo, M. Klein, A. von der Linden, M. McDonald, D. Rapetti,

C. L. Reichardt, K. Sharon, B. Stalder, S. A. Stanford, A. A. Stark, C. Stern, and A. Zenteno, "Galaxy kinematics and mass calibration in massive SZE-selected galaxy clusters to z $= 1.3$," *Monthly Notices of the Royal Astronomical Society*, vol. 482, pp. 1043–1061, 10 2018.

5 A. Vikhlinin, B. R. McNamara, W. Forman, C. Jones, H. Quintana, and A. Hornstrup, "A catalog of 203 galaxy clusters serendipitously detected in theROSATPSPC pointed observations," *The Astrophysical Journal*, vol. 502, pp. 558–581, aug 1998.

6 M. Roos, *Introduction to Cosmology*. Wiley, 4th ed., 2015.

7 G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright, "NINE-YEARWILKINSON MICROWAVE ANISOTROPY PROBE(WMAP) OBSERVATIONS: COSMOLOGICAL PARAMETER RESULTS," *The Astrophysical Journal Supplement Series*, vol. 208, p. 19, sep 2013.

8 L. Sparke and J. Gallagher, *Galaxies in the Universe*. Campbridge University Press, 2nd ed., 2007.

9 J. Sohn, M. J. Geller, A. Diaferio, and K. J. Rines, "Velocity dispersions of brightest cluster galaxies and their host clusters," 2019.

10 N. Clerc, A. Merloni, Y.-Y. Zhang, A. Finoguenov, T. Dwelly, K. Nandra, C. Collins, K. Dawson, J.-P. Kneib, E. Rozo, E. Rykoff, T. Sadibekova, J. Brownstein, Y.-T. Lin, J. Ridl, M. Salvato, A. Schwope, M. Steinmetz, H.-J. Seo, and J. Tinker, "SPIDERS: the spectroscopic follow-up of X-ray-selected clusters of galaxies in SDSS-IV," *Monthly Notices of the Royal Astronomical Society*, vol. 463, pp. 4490–4515, 09 2016.

11 T. C. Beers, K. Flynn, and K. Gebhardt, "Measures of location and scale for velocities in clusters of galaxies - A robust approach," *Astrophysical journal*, vol. 100, pp. 32–46, July 1990.

12 H. K. C. Yee and O. López-Cruz, "A quantitative measure of the richness of galaxy clusters," *The Astronomical Journal*, vol. 117, p. 1985–1994, May 1999.

13 K. Rines, M. J. Geller, A. Diaferio, and M. J. Kurtz, "Measuring the ultimate halo mass of galaxy clusters: Redshifts and mass profiles from the hectospec cluster survey (hecs)," *The Astrophysical Journal*, vol. 767, p. 15, Mar 2013.

14 Y. Rephaeli, "Comptonization of the cosmic microwave background: the sunyaevzeldovich effect," *Annual review of astronomy and astrophysics*, vol. 33, no. 1, pp. 541–579, 1995.

15 SDSS collaboration, ""http://classic.sdss.org/dr4/algorithms/spectemplates/index.html, viewed 22nd Jan 2020"."

16 M. Hart, "Long-term spectroscopic observations of the atmospheric airglow by the sloan digital sky survey," *Publications of the Astronomical Society of the Pacific*, vol. 131, p. 015003, 01 2019.

17 D. E. Osterbrock, J. P. Fulbright, A. R. Martel, M. J. Keane, S. C. Trager, and G. Basri, "Night-sky high-resolution spectral atlas of OH and o2 emission lines for echelle spectrograph wavelength calibration," *Publications of the Astronomical Society of the Pacific*, vol. 108, p. 277, mar 1996.

18 J. S. Almeida, R. Terlevich, E. Terlevich, R. C. Fernandes, and A. B. Morales-Luis, "QUALITATIVE INTERPRETATION OF GALAXY SPECTRA," *The Astrophysical Journal*, vol. 756, p. 163, aug 2012.

19 S. Shelyag, M. Schüssler, S. K. Solanki, S. V. Berdyugina, and A. Vögler, "G-band spectral synthesis and diagnostics of simulated solar magneto-convection," , vol. 427, pp. 335–343, Nov 2004.

20 J. E. Gunn, M. Carr, C. Rockosi, M. Sekiguchi, K. Berry, B. Elms, E. de Haas, Ž. Ivezić, G. Knapp, R. Lupton, G. Pauls, R. Simcoe, R. Hirsch, D. Sanford, S. Wang, D. York, F. Harris, J. Annis, L. Bartozek, W. Boroski, J. Bakken, M. Haldeman, S. Kent, S. Holm, D. Holmgren, D. Petravick, A. Prosapio, R. Rechenmacher, M. Doi, M. Fukugita, K. Shimasaku, N. Okada, C. Hull, W. Siegmund, E. Mannery, M. Blouke, D. Heidtman, D. Schneider, R. Lucinio, and J. Brinkman, "The sloan digital sky survey photometric camera," *The Astronomical Journal*, vol. 116, pp. 3040–3081, dec 1998.

21 D. J. Bates, R. Tojeiro, J. A. Newman, V. Gonzalez-Perez, J. Comparat, D. P. Schneider, M. Lima, and A. Streblyanska, "Mass functions, luminosity functions, and completeness measurements from clustering redshifts," *Monthly Notices of the Royal Astronomical Society*, vol. 486, p. 3059–3077, Apr 2019.

22 C. Hennig, J. J. Mohr, A. Zenteno, S. Desai, J. P. Dietrich, S. Bocquet, V. Strazzullo, A. Saro, T. M. C. Abbott, F. B. Abdalla, M. Bayliss, A. Benoit-Lévy, R. A. Bernstein, E. Bertin, D. Brooks, R. Capasso, D. Capozzi, A. Carnero, M. Carrasco Kind, J. Carretero, I. Chiu, C. B. D'Andrea, L. N. daCosta, H. T. Diehl, P. Doel, T. F. Eifler, A. E. Evrard, A. Fausti-Neto, P. Fosalba, J. Frieman, C. Gangkofner, A. Gonzalez, D. Gruen, R. A. Gruendl, N. Gupta, G. Gutierrez, K. Honscheid, J. Hlavacek-Larrondo, D. J. James, Kuehn, N. Kuropatkin, O. Lahav, M. March, J. L. Marshall, P. Martini, M. McDonald, P. Melchior, C. J. Miller, R. Miquel, E. Neilsen, B. Nord, R. Ogando, A. A. Plazas, C. Reichardt, A. K. Romer, E. Rozo, E. S. Rykoff, E. Sanchez, B. Santiago, M. Schubnell, I. Sevilla-Noarbe, R. C. Smith, M. Soares-Santos, F. Sobreira, B. Stalder, S. A. Stanford, E. Suchyta, M. E. C. Swanson, Tarle, D. Thomas, V. Vikram, A. R. Walker, and Y. Zhang, "Galaxy populations in massive galaxy clusters

to z = 1.1: colour distribution, concentration, halo occupation number and red sequence fraction," *Monthly Notices of the Royal Astronomical Society*, vol. 467, pp. 4015–4035, 01 2017.

23 SDSS collaboration, ""https://www.sdss.org/surveys/eboss/, viewed 21st Aug 2019"."

24 R. Ahumada, C. A. Prieto, A. Almeida, F. Anders, S. F. Anderson, B. H. Andrews, B. Anguiano, R. Arcodia, E. Armengaud, M. Aubert, S. Avila, V. Avila-Reese, C. Badenes, C. Balland, K. Barger, J. K. Barrera-Ballesteros, S. Basu, J. Bautista, R. L. Beaton, T. C. Beers, B. I. T. Benavides, C. F. Bender, M. Bernardi, M. Bershady, F. Beutler, C. M. Bidin, J. Bird, D. Bizyaev, G. A. Blanc, M. R. Blanton, M. Boquien, J. Borissova, J. Bovy, W. N. Brandt, J. Brinkmann, J. R. Brownstein, K. Bundy, M. Bureau, A. Burgasser, E. Burtin, M. Cano-Diaz, R. Capasso, M. Cappellari, R. Carrera, S. Chabanier, W. Chaplin, M. Chapman, B. Cherinka, C. Chiappini, P. D. Choi, S. D. Chojnowski, H. Chung, N. Clerc, D. Coffey, J. M. Comerford, J. Comparat, L. da Costa, M.-C. Cousinou, K. Covey, J. D. Crane, K. Cunha, G. da Silva Ilha, Y. S. Dai, S. B. Damsted, J. Darling, D. H. Darrington, J. W. D. Jr., R. Davies, K. Dawson, N. De, A. de la Macorra, N. D. Lee, A. B. de Andrade Queiroz, A. D. Machado, S. de la Torre, F. Dell'Agli, H. du Mas des Bourboux, A. M. Diamond-Stanic, S. Dillon, J. Donor, N. Drory, C. Duckworth, T. Dwelly, G. Ebelke, S. Eftekharzadeh, A. D. Eigenbrot, Y. P. Elsworth, M. Eracleous, G. Erfanianfar, S. Escoffier, X. Fan, E. Farr, J. G. Fernandez-Trincado, D. Feuillet, A. Finoguenov, P. Fofie, A. Fraser-McKelvie, P. M. Frinchaboy, S. Fromenteau, H. Fu, L. Galbany, R. A. Garcia, D. A. Garcia-Hernandez, L. A. G. Oehmichen, J. Ge, M. A. G. Maia, D. Geisler, J. Gelfand, J. Goddy, J.-M. L. Goff, V. Gonzalez-Perez, K. Grabowski, P. Green, C. J. Grier, H. Guo, J. Guy, P. Harding, S. Hasselquist, A. J. Hawken, C. R. Hayes, F. Hearty, S. Hekker, D. W. Hogg, J. Holtzman, J. Hou, B.-C. Hsieh, D. Huber, J. A. S. Hunt, J. I. Chitham, J. Imig, M. Jaber, C. E. J. Angel, J. A. Johnson, A. M. Jones, H. Jonsson, E. Jullo, Y. Kim, K. Kinemuchi, C. C. K. IV, G. W. Kite, M. Klaene, J.-P. Kneib, J. A. Kollmeier, H. Kong, M. Kounkel, D. Krishnarao, I. Lacerna, T.-W. Lan, R. R. Lane, D. R. Law, H. W. Leung, H. Lewis, C. Li, J. Lian, L. Lin, D. Long, P. Longa-Pena, B. Lundgren, B. W. Lyke, J. T. Mackereth, C. L. MacLeod, S. R. Majewski, A. Manchado, C. Maraston, P. Martini, T. Masseron, K. L. Masters, S. Mathur, R. M. McDermid, A. Merloni, M. Merrifield, S. Meszaros, A. Miglio, D. Minniti, R. Minsley, T. Miyaji, F. G. Mohammad, B. Mosser, E.-M. Mueller, D. Muna, A. Munoz-Gutierrez, A. D. Myers, S. Nadathur, P. Nair, J. C. do Nascimento, R. J. Nevin, J. A. Newman, D. L. Nidever, C. Nitschelm, P. Noterdaeme, J. E. O'Connell, M. D. Olmstead, D. Oravetz, A. Oravetz, Y. Osorio, Z. J. Pace, N. Padilla, N. Palanque-Delabrouille, P. A. Palicio, H.-A. Pan, K. Pan, J. Parker, R. Paviot, S. Peirani, K. P. Ramrez, S. Penny, W. J. Percival, I. Perez-Fournon, I. Perez-Rafols, P. Petitjean, M. M. Pieri, M. Pinsonneault, V. J. Poovelil, J. T. Povick, A. Prakash, A. M. Price-Whelan, M. J. Raddick, A. Raichoor, A. Ray, S. B. Rembold, M. Rezaie, R. A. Riffel, R. Riffel, H.-W.

Rix, A. C. Robin, A. Roman-Lopes, C. Roman-Zuniga, B. Rose, A. J. Ross, G. Rossi, K. Rowlands, K. H. R. Rubin, M. Salvato, A. G. Sanchez, L. Sanchez-Menguiano, J. R. Sanchez-Gallego, C. Sayres, A. Schaefer, R. P. Schiavon, J. S. Schimoia, E. Schlafly, D. Schlegel, D. P. Schneider, M. Schultheis, A. Schwope, H.-J. Seo, A. Serenelli, A. Shafieloo, S. J. Shamsi, Z. Shao, S. Shen, M. Shetrone, R. Shirley, V. S. Aguirre, J. D. Simon, M. F. Skrutskie, A. Slosar, R. Smethurst, J. Sobeck, B. C. Sodi, D. Souto, D. V. Stark, K. G. Stassun, M. Steinmetz, D. Stello, J. Stermer, T. Storchi-Bergmann, A. Streblyanska, G. S. Stringfellow, A. Stutz, G. Suarez, J. Sun, M. Taghizadeh-Popp, M. S. Talbot, J. Tayar, A. R. Thakar, R. Theriault, D. Thomas, Z. C. Thomas, J. Tinker, R. Tojeiro, H. H. Toledo, C. A. Tremonti, N. W. Troup, S. Tuttle, E. Unda-Sanzana, M. Valentini, J. Vargas-Gonzalez, M. Vargas-Magana, J. A. Vazquez-Mata, M. Vivek, D. Wake, Y. Wang, B. A. Weaver, A.-M. Weijmans, V. Wild, J. C. Wilson, R. F. Wilson, N. Wolthuis, W. M. Wood-Vasey, R. Yan, M. Yang, C. Yeche, O. Zamora, P. Zarrouk, G. Zasowski, K. Zhang, C. Zhao, G. Zhao, Z. Zheng, Z. Zheng, G. Zhu, and H. Zou, "The sixteenth data release of the sloan digital sky surveys: First release from the apogee-2 southern survey and full release of eboss spectra," 2019.

25 M. D. Gladders and H. K. C. Yee, "A new method for galaxy cluster detection. i. the algorithm," *The Astronomical Journal*, vol. 120, pp. 2148–2162, oct 2000.

26 E. S. Rykoff, E. Rozo, M. T. Busha, C. E. Cunha, A. Finoguenov, A. Evrard, J. Hao, B. P. Koester, A. Leauthaud, B. Nord, M. Pierre, R. Reddick, T. Sadibekova, E. S. Sheldon, and R. H. Wechsler, "redMaPPer. I. Algorithm and SDSS DR8 Catalog," , vol. 785, p. 104, Apr 2014.

27 E. S. Rykoff, B. P. Koester, E. Rozo, J. Annis, A. E. Evrard, S. M. Hansen, J. Hao, D. E. Johnston, T. A. McKay, and R. H. Wechsler, "Robust optical richness estimation with reduced scatter," *The Astrophysical Journal*, vol. 746, p. 178, Feb 2012.

28 T. Boller, "2rxs—the deepest and cleanest x-ray all-sky catalogue before erosita," *Astronomische Nachrichten*, vol. 338, no. 9-10, pp. 972–977, 2017.

29 A. Finoguenov, E. Rykoff, N. Clerc, M. Costanzi, S. Hagstotz, J. I. Chitham, K. Kiiveri, C. C. Kirkpatrick, R. Capasso, J. Comparat, S. Damsted, R. Dupke, G. Erfanianfar, J. P. Henry, F. Kaefer, J.-P. Kneib, V. Lindholm, E. Rozo, L. van Waerbeke, and J. Weller, "Codex clusters. the survey, the catalog, and cosmology of the x-ray luminosity function," 2019.

30 N. Clerc, C. C. Kirkpatrick, A. Finoguenov, R. Capasso, J. Comparat, S. Damsted, K. Furnell, A. E. Kukkola, J. Ider Chitham, A. Merloni, A. Gueguen, T. Dwelly, C. Collins, A. Saro, G. Erfanianfar, G. Mamon, N. Padilla, E. Jullo,, "SPIDERS: overview of the X-ray galaxy cluster follow-up and the final spectroscopic data release," 2020.

31 J. Ruel, G. Bazin, M. Bayliss, M. Brodwin, R. J. Foley, B. Stalder, K. A. Aird, R. Armstrong, M. L. N. Ashby, M. Bautz, B. A. Benson, L. E. Bleem, S. Bocquet, J. E. Carlstrom, C. L. Chang, S. C. Chapman, H. M. Cho, A. Clocchiatti, T. M. Crawford, A. T. Crites, T. de Haan, S. Desai, M. A. Dobbs, J. P. Dudley, W. R. Forman, E. M. George, M. D. Gladders, A. H. Gonzalez, N. W. Halverson, N. L. Harrington, F. W. High, G. P. Holder, W. L. Holzapfel, J. D. Hrubes, C. Jones, M. Joy, R. Keisler, L. Knox, A. T. Lee, E. M. Leitch, J. Liu, M. Lueker, D. Luong-Van, A. Mantz, D. P. Marrone, M. McDonald, J. J. McMahon, J. Mehl, S. S. Meyer, L. Mocanu, J. J. Mohr, T. E. Montroy, S. S. Murray, T. Natoli, D. Nurgaliev, S. Padin, T. Plagge, C. Pryke, C. L. Reichardt, A. Rest, J. E. Ruhl, B. R. Saliwanchik, A. Saro, J. T. Sayre, K. K. Schaffer, L. Shaw, E. Shirokoff, J. Song, R. Šuhada, H. G. Spieler, S. A. Stanford, Z. Staniszewski, A. A. Starsk, K. Story, C. W. Stubbs, A. van Engelen, K. Vanderlinde, J. D. Vieira, A. Vikhlinin, R. Williamson, O. Zahn, and A. Zenteno, "OPTICAL SPECTROSCOPY AND VELOCITY DISPERSIONS OF GALAXY CLUSTERS FROM THE SPT-SZ SURVEY," *The Astrophysical Journal*, vol. 792, p. 45, aug 2014.

32 C. C. Kirkpatrick, N. Clerc, A. Finoguenov, S. Damsted, J. Ider Chitham, A. E. Kukkola, A. Gueguen, K. Furnell, A. Saro, R. Capasso, N. Padilla, G. Erfanianfar, J. Comparat, G. Mamon, C. Collins, A. Merloni, "SPIDERS: overview of the largest spectroscopically conrmed X-ray galaxy cluster catalogue," 2020.

33 A. Hou, L. C. Parker, W. E. Harris, and D. J. Wilman, "Statistical tools for classifying galaxy group dynamics," *The Astrophysical Journal*, vol. 702, p. 1199–1210, Aug 2009.

34 J. Pinkney, K. Roettiger, J. Burns, and C. Bird, "Evaluation of statistical tests for substructure in clusters of galaxies," *The Astrophysical Journal Supplement Series*, vol. 104, p. 1, 04 1996.

35 A. Hou, L. Parker, D. Wilman, S. McGee, W. Harris, J. Connelly, M. Balogh, J. Mulchaey, and R. Bower, "Substructure in the most massive geec groups : field-like populations in dynamically active groups.," *Monthly notices of the Royal Astronomical Society.*, vol. 421, pp. 3594–3611, April 2012.

36 A. Dressler and S. Shectman, "Evidence for substructure in rich clusters of galaxies from radial-velocity measurements," *Astron. J.; (United States)*, vol. 95, 4 1988.